

EDITED BY

MICHAEL
GLANZBERG



≡ The Oxford Handbook of
TRUTH

THE OXFORD HANDBOOK OF

TRUTH

THE OXFORD HANDBOOK OF

TRUTH

Edited by

MICHAEL GLANZBERG

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© the several contributors 2018

The moral rights of the authors have been asserted

First Edition published in 2018

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2017959372

ISBN 978-0-19-955792-9

Printed and bound by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

ACKNOWLEDGMENTS

This volume was very long in the making. My main debt of gratitude is to the authors who have contributed, and to Peter Momtchiloff at Oxford University Press. All of them offered generous support and incredible patience through the long process of bringing this volume to fruition. I benefited frequently from the advice of many collaborators, friends, and teachers, including Jc Beall, the late George Boolos, Christopher Gauker, Warren Goldfarb, Volker Halbach, Richard Kimberly Heck, Jeffrey C. King, Ernie Lepore, Charles Parsons, Agustín Rayo, Sanford Shieh, Zoltán Gendler Szabó, and Gabriel Uzquiano. Thanks do not express what I owe to Ching-I. Her strength and her grace amaze and inspire me. Three friends who would have found it interesting passed away while this volume was being prepared: Aldo Antonelli, Delia Graff Fara, and James Higginbotham. This volume is dedicated to their memory.

CONTENTS

<i>List of contributors</i>	xi
-----------------------------	----

Introduction	1
MICHAEL GLANZBERG	

PART I ANCIENT AND MODERN THEORIES OF TRUTH

1. Plato and Aristotle on Truth and Falsehood	9
JAN SZAIF	
2. Truth in the Middle Ages	50
MARGARET CAMERON	
3. Early Modern Theories of Truth	75
ALAN NELSON	
4. Idealism and the Question of Truth	93
CLINTON TOLLEY	

PART II TRUTH IN EARLY ANALYTIC PHILOSOPHY

5. Truth in British Idealism and its Analytic Critics	125
THOMAS BALDWIN	
6. Judgments, Facts, and Propositions: Theories of Truth in Russell, Wittgenstein, and Ramsey	150
PETER SULLIVAN AND COLIN JOHNSTON	
7. Truth in Frege	193
RICHARD KIMBERLY HECK AND ROBERT MAY	

PART III THE CLASSICAL THEORIES OF TRUTH

- 8. The Coherence Theory of Truth 219
RALPH C. S. WALKER
- 9. The Correspondence Theory of Truth 238
MARIAN DAVID
- 10. The Identity Theory of Truth 259
STEWART CANDLISH AND NIC DAMNJANOVIC
- 11. The Pragmatist Theory of Truth 283
CHERYL MISAK

PART IV TRUTH IN METAPHYSICS

- 12. Propositions and Truth-Bearers 307
JEFFREY C. KING
- 13. Truthmakers 333
ROSS P. CAMERON
- 14. A Logical Theory of Truth-Makers and Falsity-Makers 355
NEIL TENNANT
- 15. Bivalence and Determinacy 394
IAN RUMFITT
- 16. Truth, Objectivity, and Realism 433
SANFORD SHIEH
- 17. Deflationist Truth 477
JODY AZZOUNI
- 18. Truth in Fictionalism 503
ALEXIS BURGESS
- 19. Relative Truth 517
HERMAN CAPPELEN AND TORFINN THOMESSEN HUVENES
- 20. Truth Pluralism 543
NIKOLAJ J. L. L. PEDERSEN AND MICHAEL P. LYNCH

PART V OTHER APPLICATIONS

- | | |
|-----------------------------|-----|
| 21. The Moral Truth | 579 |
| MARK SCHROEDER | |
| 22. Truth and the Sciences | 602 |
| ANJAN CHAKRAVARTY | |
| 23. Truth and Truthlikeness | 625 |
| GRAHAM ODDIE | |
| 24. Truth in Mathematics | 648 |
| ØYSTEIN LINNEBO | |

PART VI FORMAL THEORIES AND PARADOX

- | | |
|--|-----|
| 25. Semantic Paradoxes: A Psychohistory of Self-Defeat | 669 |
| ROY A. SORENSEN | |
| 26. Tarski on the Concept of Truth | 695 |
| GREG RAY | |
| 27. The Axiomatic Approach to Truth | 718 |
| KENTARO FUJIMOTO AND VOLKER HALBACH | |
| 28. Non-Classical Theories of Truth | 739 |
| JC BEALL AND DAVID RIPLEY | |
| 29. Contextual Theories of Truth and Paradox | 755 |
| KEITH SIMMONS | |
| <i>Index</i> | 787 |

LIST OF CONTRIBUTORS

Jody Azzouni, Professor of Philosophy, Tufts University

Thomas Baldwin, Professor of Philosophy Emeritus, University of York, Emeritus

Jc Beall, Board of Trustees Distinguished Professor of Philosophy at the University of Connecticut; Professor of Philosophy, University of Tasmania.

Alexis Burgess, independent scholar

Margaret Cameron, Canada Research Council Chair in the Aristotelian Tradition and Associate Professor of Philosophy, University of Victoria

Ross P. Cameron, Professor of Philosophy, University of Virginia

Stewart Candlish, Senior Honorary Research Fellow, The University of Western Australia

Herman Cappelen, Professor of Philosophy, University of Oslo and University of St Andrews

Anjan Chakravartty, Appignani Foundation Chair and Professor of Philosophy, University of Miami

Nic Damnjanovic, Barrister, Fountain Court Chambers, London; formerly Assistant Professor of Philosophy, The University of Western Australia

Marian David, University Professor, Karl-Franzens University of Graz

Kentaro Fujimoto, Lecturer in Mathematics and Philosophical Logic, University of Bristol

Volker Halbach, Professor of Philosophy, Oxford University and Fellow of New College, Oxford

Richard Kimberly Heck, Professor of Philosophy, Brown University

Torfinn Thomesen Huvenes, Lecturer in the Department of Historical, Philosophical and Religious Studies, Umeå University

Colin Johnston, Senior Lecturer in Philosophy, University of Stirling

Jeffrey C. King, Distinguished Professor of Philosophy, Rutgers University

Øystein Linnebo, Professor of Philosophy, University of Oslo

Michael P. Lynch, Professor of Philosophy, University of Connecticut

Robert May, Distinguished Professor of Philosophy, University of California, Davis

Cheryl Misak, University Professor and Professor of Philosophy, University of Toronto

Alan Nelson, Professor of Philosophy, University of North Carolina at Chapel Hill

Graham Oddie, Professor of Philosophy, University of Colorado, Boulder

Nikolaj J. L. L. Pedersen, Associate Professor of Philosophy, Yonsei University

Greg Ray, Associate Professor of Philosophy, University of Florida

David Ripley, Lecturer in Philosophy, Monash University

Ian Rumfitt, Professor of Philosophy, Oxford University and Senior Research Fellow of All Souls College, Oxford

Mark Schroeder, Professor of Philosophy, University of Southern California

Sanford Shieh, Associate Professor of Philosophy, Wesleyan University

Keith Simmons, Associate Professor of Philosophy, University of Connecticut

Roy A. Sorensen, Professor of Philosophy, Washington University, St Louis

Peter Sullivan, Professor of Philosophy, University of Stirling

Jan Szaif, Professor of Philosophy, University of California, Davis

Neil Tennant, Arts and Sciences Distinguished Professor in Philosophy, Ohio State University

Clinton Tolley, Associate Professor of Philosophy, University of California, San Diego

Ralph C. S. Walker, Emeritus Fellow in Philosophy of Magdalen College, Oxford

INTRODUCTION

MICHAEL GLANZBERG

READERS hold before them a volume bearing the title *Handbook of Truth*. A title like this suggests something very demanding: a book to keep at hand in which to find truths. Such a book would be useful indeed! But we might imagine it to be so long as to be unwieldy. After all, such a book would in effect have to list all truths, or at least enough of them to be worthy of the title “Handbook.”

But there is no such book, and I doubt there could ever be one. For all the folly involved in editing a large Handbook, I certainly did not set out to edit *that* book. But care about wording reminds us the book that does not exist would have been better called the *Handbook of Truths*. The book presented here is rather a *Handbook of Truth*. Its topic is not all truths, enumerated for convenient use, but simply truth.

In some ways, the topic of this Handbook is easy state. Truth is a property, that has been of great interest to philosophy and beyond for about as long as we have a history for philosophy. It is a topic of active research in many branches of philosophy, notably metaphysics and logic. That property is the topic of this Handbook. Essays in this volume will explore the basic nature of the property of truth, i.e. its underlying metaphysical and logical properties. Others explore its role in such areas as ethics and science and mathematics, and so on. They do so from a number of perspectives, both historical and systematic. They use various techniques and approaches, from metaphysics, ethics, philosophy of science, logic, and many others.

Experts will recognize, of course, that most any claim about truth is controversial, including ones I presupposed in stating the topic of this Handbook. Whether truth is really a substantial property with any underlying metaphysics, whether there are interesting logical questions about truth, and so on, are all points that have been challenged over the years. I am reminded of the way J. L. Austin started his classic paper “Truth:”¹

“What is truth?” said jesting Pilate, and would not stay for an answer. Pilate was in advance of his time. For “truth” itself is an abstract noun, a camel, that is, of a logical

¹ J. L. Austin (1950), “Truth,” *Proceedings of the Aristotelian Society Supplementary Volumes* 24: 111–28.

construction, which cannot get past the eye even of a grammarian. We approach it cap and categories in hand: we ask ourselves whether Truth is a substance (the Truth, the Body of Knowledge), or a quality (something like the colour red, inhering in truths), or a relation (“correspondence”). But philosophers should take something more nearly their own size to strain at. What needs discussing rather is the use, or certain uses, of the word “true.” *In vino*, possibly, “*veritas*,” but in sober symposium “*verum*.”

Philosophers have indeed approached truth with categories, and maybe caps, in hand (its status as a camel I shall not guess at); and indeed some philosophers have seen some questions about the nature of truth to be too much to strain at, and have urged us to cut them down to a smaller size.

A Handbook of truth seeks to review these many sorts of issues about truth itself: what, if any, metaphysical or logical nature the property of truth has, what its connections with other areas of inquiry might be, how it has been addressed over the years, and what theories have been developed. All of these topics are represented in this Handbook. I shall not try to summarize each essay the Handbook offers. Instead, I shall offer an overview of the themes that it addresses, and how the essays in it contribute to them.

1 HISTORICAL THEMES

The study of truth is old. How old is hard to say, but we know the presocratics had a great deal to say about truth (as even casual readers of Plato will see). So it is old indeed, and for all we know, may have started with the dawn of philosophical thinking. It is inviting to speculate about how thinking about core philosophical topics like the nature of truth arose so long ago, but we do at least have some evidence that they did.

From as far back as we have records, truth has from time to time been a topic of intense philosophical investigation. Important thinking about truth occurred in the medieval period, for instance. Metaphysics and logic were both lively topics in that time, and the study of truth flourished. In a somewhat different way, it also flourished in the post-Kantian period, though with concerns about idealism more in the forefront than logic. Not every historical period has seen truth as a primary philosophical subject, however. It is often said that the early modern period did not focus much on the theory of truth, as epistemology was of paramount concern.

This Handbook offers a few samplings of these historical issues, though a detailed study of all of them is beyond what any one volume can do. There are two Parts devoted specifically to the history of the study of truth; though of course, many of the essays in the volume address historical issues. The first Part, entitled “Ancient and Modern Theories of Truth,” addresses the earlier history. An essay here reviews central ideas of Plato and Aristotle on truth, which will give readers an introduction to how truth was approached in the ancient period. Some of the many topics in medieval theories of truth

are also surveyed, along with some of their metaphysical, epistemological, and theological implications. The early modern period is also reviewed, and you will find that the common view that there is little said about truth in the early modern period is not the full story. Some themes in the discussion of truth in the post-Kantian period are also explored, with an emphasis on their connection to idealism. These samplings will, I hope, give readers an overview of some of the fascinating issues in the history of the study of truth, and some insight into some of the important moments in that history.

Part II, “Truth in Early Analytic Philosophy,” delves more deeply into one historical period where truth was an extremely active subject: the beginnings of what is often called “analytic philosophy.” This period, around the dawn of the twentieth century, sees important work of the founding figures of the analytic tradition, such as Russell, Moore, Frege, Wittgenstein, and Ramsey, and of course their opponents in the British idealist tradition such as Bradley. Truth is crucial to the disagreement between Bradley, espousing a form of idealist monism, and Russell and Moore, who espouse a sequence of distinct views of truth, facts, and related issues in rejecting this idealist tradition. One essay in this Part explores the dispute between Bradley and Russell and Moore, while another explores the many ideas about truth, facts, and propositions that especially animated the early work of Russell, followed by Wittgenstein and Ramsey. Frege’s work is also central to the origins of the analytic tradition, though he worked somewhat outside the British tradition, steeped in its own distinctive form of idealism at the end of the nineteenth century. A mathematician by training, he was certainly familiar with the post-Kantian tradition that dominated German philosophy, but approached many philosophical issues in a novel way. Frege’s views on truth have been both influential and controversial. They are discussed in the final essay of Part II.

2 CLASSICAL THEORIES

The emergence of analytic philosophy at the beginning of the twentieth century, and Russell and Moore’s turn away from British idealism, also led to the development of a range of theories of truth that have been important in the ensuing years. Together with theories that emerged from work in the American pragmatist tradition, they offer what I have labeled the “classical” theories of truth: the identity, correspondence, coherence, and pragmatist theories. These theories have historical roots in the turn of the twentieth century, but they have taken on lives of their own, and have become topics in their own right. Each of these is discussed by a separate essay in Part III, entitled “The Classical Theories of Truth.” The theories are classical in that they are starting points for a number of subsequent investigations, and offer an interesting map of the landscape of theories of truth. Of course, such classical theories are often modified over time, and they can be formulated in many ways. The essays in Part III review a number of options for formulating the classical theories. In spite of the many changes such theories have undergone, a great deal of thinking in recent years and before has been strongly influenced by them.

3 THEMES IN THE METAPHYSICS AND APPLICATIONS OF TRUTH

Parts IV and V move away from concerns with the history and the classical theories, to a range of ways to approach truth and the many issues connected closely to truth. One of the closest connections we find is between truth and many issues in metaphysics. A number of these are reviewed in Part IV, entitled “Truth in Metaphysics.” As with the historical sections, the range of issues and theories that have been discussed over the years connecting truth and metaphysics is huge, and impossible to survey comprehensively. Again, the essays in this Part offer some samplings of important issues, theories, and techniques for exploring the metaphysics of truth. Several explore the apparatus we need to describe truth, or some of its main metaphysical consequences. The property of truth applies to something, and there is a rich tradition of thinking these truth-bearers—often identified as propositions—are an important area of study in their own right. Likewise, in metaphysics, there has been a lively discussion of what are called “truth-makers”: usually (thick) particulars that can witness the truth of propositions. Truth-makers play a special role in certain metaphysical theories, though they obviously are of interest to the theory of truth itself. Both these topics are explored in essays in this Part IV.

When we come to consequences of the theory of truth, one important set of issues relates to objectivity, determinacy, and related matters. These have been of special importance to many discussions of truth since the 1950s, though of course, they were also of interest to Aristotle, and are perennial topics of philosophical concern. Two essays provide samplings of these issues. One explores the relation of truth to realism and objectivity, focusing on the discussion beginning in the 1950s following important work by Michael Dummett. The other takes up a range of questions about how we might understand determinism, and its relation to truth. All these issues relate closely to issues in logic, as was made clear by developments in intuitionistic logic. A logical approach to them is presented in one of the essays in Part IV.

One of the most striking more recent developments in the metaphysics of truth is a more “deflationary” view of the property of truth itself. All the classical theories see truth as a substantial property of some kind, though they differ widely on just what kind, and how much it is a metaphysical, epistemological, or practical property. In contrast, many recent approaches doubt truth is so substantial. These doubts have taken many forms over the years. Two essays provide samplings of how such doubts may be formulated, and what kinds of theory may result from them. One takes up the much-discussed theme of deflationism about truth, that offers a view that truth is not the metaphysically or epistemologically substantial concept we might have expected. A parallel thought in recent years has been that many metaphysically problematic areas are really more like fictions. How this might be applied to theories of truth is explored in another essay in Part IV.

Another important question, that has been of perennial interest but a focus of research recently, is that of whether we can make sense of truth as a relative property. Such a relativist property of truth might, for instance, make something true for you or true for me, rather than just true. Recent ideas about relative truth are explored in a further essay in this Part. Finally, a topic that has emerged in recent years pertains to whether there is a single notion of truth, or whether there are distinct properties that apply to different domains. Perhaps moral truth is a different property from logical truth, to illustrate. If so, what are those properties, and what makes them all really properties of truth? This last topic concludes Part IV.

Truth connects to a very wide range of issues. Many of them relate to other areas of inquiry that have been of special concern to philosophy. Often these questions can be put succinctly in terms of truth. Is some area of inquiry, like say ethics, directed toward truth, or are the statements of such an area true or false? Such questions are not restricted to ethics, of course. Philosophers have raised them frequently about mathematics. Perhaps asking the questions in just that form about science could have the ring of skepticism. But there are many non-skeptical questions to ask about where and how science works toward truth, and what to make of well-confirmed statements of science which we know could be revised as our knowledge grows. Such statements would, in such a case, not be true, but would stand in an important relation to truth. Part V, entitled “Other Applications,” offers some samplings of the many issues that have been explored related to questions like this. One essay examines the much-discussed question of what truthlikeness might be, and how it relates to truth. The general questions of the nature of truth in science, along with ethics and mathematics, are taken up by individual essays.

4 LOGICAL THEMES

The final Part of this Handbook is most closely tied to logic. A glance at the historical sections will show that matters of logic and of truth have gone together throughout their history, and were extremely important in the medieval and early analytic periods. They have also been central to an important strand of contemporary research on theories of truth (one that is near and dear to the heart of the editor of this Handbook). As has been known since ancient times, the liar paradox makes trouble for what seem to be simple and obvious principles governing reasoning with truth—principles so simple and obvious that they are candidates to be logical principles. In light of constructions like “This sentence is not true,” those simple and obvious principles threaten to be logically incoherent. Discovering the correct principles of reasoning about truth—perhaps the logic of truth—turns out to be a complicated matter because of this. In recent years, it has been a focus for a range of important questions not just about truth, but about logic, and has engaged an important collection of ideas and techniques of modern logic.

Part VI, entitled “Formal Theories and Paradox,” offers a sampling of these issues. The nature of the liar and related paradoxes, and what adequate responses to them might be

like, is reviewed, with an eye toward the long, rich history of this topic. In current times, much of the work on formal theories of truth and paradox has taken the seminal work of Tarski as its starting place (though often to reject Tarski's conclusions). Tarski's ideas are explored in a specific essay. The major approaches to truth and paradox in the current literature can be broken down into two main groups. One concludes that the simple principles governing truth are corrected, and so in some ways logic cannot be classical in order for logical coherence to be retained. (This is sometimes, but not always, put in terms of revising classical logic.) Note, a striking claim about the nature of logic is being offered here! The alternative is to retain classical logic, but modify something about the principles governing truth. Much recent work on this approach has been done by developing mathematically rich axiomatic theories of truth; while other work has sought to invoke ideas from the philosophy of language to explain how the principles of reasoning for truth work, and how they can avoid paradox. Again, each of these themes is explored by a separate essay.

5 INVITATION

As I have mentioned several times, the history of the study of truth is long; the range and variety of issues it relates to, both within and outside of philosophy, is immense; and the problems it raises for metaphysics, logic, and many more areas within philosophy are among the most difficult and persistent ones we know. Truth is a big and hard topic. As I have also mentioned several times, even a Handbook of the length of this one cannot hope to cover every topic related to truth. I have several times described what this volume offers as "samplings." I believe they are good representatives of the range of issues and problems truth presents, and I hope that, taken together, they offer the reader a good indication of some of the important landmarks in the study of truth. It is easy to list things left out of this Handbook. Some were omitted simply because of the vicissitudes of collecting the material for a volume of this length. But some are missing because editorial decisions had to be made to create a work that could be bound as one volume.

With those apologies, I offer the readers the twenty-nine essays in this volume, the hard work of thirty-six authors. As a Handbook, I hope it serves as a useful reference for researchers working on truth, or those new to the topic. As an overview of a large and complex topic, I hope it serves as a useful map of difficult but fascinating terrain.

PART I

ANCIENT AND
MODERN THEORIES
OF TRUTH

CHAPTER 1

PLATO AND ARISTOTLE ON TRUTH AND FALSEHOOD

JAN SZAIF

1.1 THE PRE-PHILOSOPHICAL USE OF TRUTH TERMINOLOGY

MUCH of our established philosophical terminology is the result of the creation of new words or the giving of new meanings to existing words.¹ This phenomenon is, to some extent, concealed by the fact that many of those neologisms, or new meanings, have gradually been absorbed into the colloquial language (as with words like “object,” “perception,” or “reality”). Yet there are also those cases in which philosophical theory focuses on an indispensable element of the basic conceptual framework of any language, and the study of the concept of truth should be one such case. For truth vocabulary relates to a semantic aspect of language that is so foundational that it can’t be absent from any fully developed human language.

Contemporary theories of truth typically try to analyze the semantic basis of our pre-philosophical notion of truth. Given the foundational role of truth vocabulary across languages, we would therefore expect discussions regarding the concept of truth in ancient Greek philosophy to be easily comparable to contemporary analytical discussions. Yet a closer look at the use of truth terminology in ancient Greek theories reveals that we can’t take it for granted, without further argument, that “true” has essentially the same meaning as its Greek counterpart *alēthēs*. First, it is not always easy to translate the adjective *alēthēs*, or the corresponding noun *alētheia*, as “true” or “truth” respectively. Secondly, when we look, for instance, at the passages in which Plato, one of the

¹ This chapter supersedes my earlier survey in Enders and Szaif (2006). I want to thank the editor, Michael Glanzberg, and David Charles, Mitzi Lee, John Malcolm, Adam Sennet, and Robert May for their valuable feedback.

major Greek philosophers, addresses the notion of truth (or *alētheia*), it is only in some of them that he discusses propositional truth (i.e. truth as a property of propositionally structured entities), while in others he appears to be talking about something closer to our notions of reality and genuineness. Such (from our point of view) unusual ways of talking about truth might be due to the fact that already the pre-philosophical usage of the word *alēthēs* partly deviates from English usage, or they might reflect theoretical interests and developments that led to a disconnection of the philosophical notion from the truth concept anchored in colloquial language. In order to get a more accurate perspective on the philosophical development of the concept of truth in Plato and Aristotle, we should begin with a discussion of the pre-philosophical usage of truth terminology in Greek.

Let me first comment briefly on a thesis that used to be rather popular among students of ancient Greek philosophy in Europe. According to this approach (which was not started but later often influenced by Heidegger),² the “original” Greek concept of truth is quite different from our modern understanding of truth due to the influence of the etymology of the Greek word for truth on Greek thinking. The most probable etymological explanation links *alētheia* to the stem *lēth/lath* (being concealed/escaping notice; forgetfulness) and the negative prefix *a-* (*α-privativum*). Even if we can’t say with ultimate certainty that this etymology is correct, at least it is clear that for native speakers of Greek it was the most obvious derivation (as has been documented by Heitsch 1962). Philological attempts at analyzing the “original” Greek notion of truth have focused on the use of the term *alētheia* in the early Greek epic, especially Homer, since this represents the earliest stratum of Greek literature. Thanks to these investigations we know that in Homeric literature this expression is applied much more narrowly and in a way that shows a certain closeness to its likely etymological meaning “what is unconcealed.”³ Yet this does not justify the conclusion that the notion of plain propositional truth and falsehood was still alien to the “original” Greek understanding of truth. The early Greek epic has a very rich vocabulary of truth, and *alētheiē* is only one among many alternative expressions (cf. the comprehensive survey in Luther 1935). Particular attention, in my opinion, should be given to the cognates *eteos*, *etymos*, and *etētymos*, which in Homer are virtually synonymous and exhibit the same array of basic meanings as can be found in the later usage of *alēthēs* in classical Greek, including straightforward propositional truth.⁴ I submit that the word *alēthēs* in classical (Attic) Greek has taken over the semantic roles of the adjectives of the *eteos*-group, which are rarely used in classical

² Cf. Szaif (1996: 145ffn. 92). There has also been a theological strand in this debate based on a comparison of the ancient Greek and Hebrew notions of truth (e.g. Rudolf Bultmann, Hans von Soden).

³ It is typically used in situations of *account-giving*, and the underlying idea can be rendered, in our modern vocabulary, as follows: the *alētheiē* to be recounted are the *facts as they present themselves (to an account receiver) if nothing about them is concealed (by an account giver)*; cf. Boeder (1959); Krischer (1965); Mourelatos (2008: 63f). Examples for this usage in Homer: *Iliad* 6.382, 23.361, 24.407; *Odyssey* 3.247, 3.254, 7.297, 11.507, 13.254, 14.125, 16.61, 16.226, 17.15, 17.108, 17.122, 18.342, 21.212, 22.420.

⁴ See for instance *Iliad* 2.300, 10.534; *Odyssey* 19.203 (cf. Hesiod, *Theog.* 27f), 19.567.

Greek. The peculiar etymology of *alētheiē* should therefore not be of much concern for the interpretation of the Greek understanding of truth in general.

My next step is to give a brief overview of the usages of *alēthēs/alētheia* in classical Greek based on my own systematic survey of the evidence provided by the Platonic dialogues (Szaif 1996: 25–71 and 533–5). Plato's works are not only of great significance for our understanding of the development of the concept of truth in Greek philosophy but also present us, thanks to their dialogical character, with an exceptionally broad sample of non-technical uses of the truth terminology. The most important result of such a survey is the centrality of usages expressing propositional truth ascriptions, which disproves the claim that propositional truth was not central for the pre-theoretical Greek understanding of truth. This cannot be argued here in detail. I will simply list, and briefly comment on, the main types of usage.

1.1.1 Propositional truth

While the use of *alēthēs* as an adjective determining a noun like “statement” or “belief” is statistically rather rare in Plato, propositional truth is very often expressed by a type of locution in which the form *alēthē* (neuter plural) fills the object slot of the verb: *alēthē legein* literally: “saying true [stuff]” or “saying something that is true” (e.g. *Euphr.* 8d3, *Phd.* 65c1, and *passim*). It closely corresponds to the construction “What you are saying is true” (*alēthē esti ha legeis*, e.g. *Apol.* 40e5, *Phd.* 70b1).

Even though the locutions just mentioned are clearly concerned with propositional truth plain and simple, the fact that the truth word is placed in the object position has some significance. Once philosophical reflection about propositional truth gets started, this grammatical construction can have the effect that the propositionally structured entities which function as the truth-bearers (i.e. the “true stuff” that gets asserted) are viewed as the counterparts of the linguistic entities and as belonging on the world-side of the world/world divide. Of course, the “true stuff” in question might also be identified with the sequence of words uttered (type or token), or their abstract propositional content. But we will see that there is a tendency, at least until Aristotle, to identify that which is asserted with the world-correlate of the assertion (especially in the case of affirmative assertions).

This tendency is reinforced by the fact that *alēthē* in the object position can be replaced, without change of meaning, with the participle form of “to be” (e.g. *Tht.* 179c7, *Rep.* 389c): “something that is” (*onta*). This phenomenon is an example of a specific use of the Greek verb for “to be” (*einai*) that requires the translation “is the case/is true” and applies to a propositionally structured referent (for instance the content of an assertion: “*This* [i.e. what you have just said] *is* [i.e. is the case]”—*esti tauta*, e.g. *Phd.* 65c10). Following Charles Kahn, we can call this its *veridical* use and speak of “veridical being” (sometimes also called “alethic being” or “veritative being”). *Pace* Kahn,⁵

⁵ Charles Kahn has suggested that there is a clear grammatical distinction between the veridical and the various existential uses of “to be” (*einai*) because the former includes a comparative construction

however, the distinction between the veridical and the existential uses often becomes blurred, not least because, in Greek thinking, facts involving temporal situations are viewed as themselves temporal. Such facts come to be and cease to be as the objects change in the relevant respects. This allows for gradual transitions between temporal facts, situations, chains of events, constellations of objects, and, finally, the existence of objects.⁶ The lack of a clear boundary between existential and veridical uses contributes, as we will see, to certain puzzles about falsehood in Greek philosophy.

As long as the speaker focuses on truth or falsity of a statement or belief, no article is added in front of *alēthē*. If, however, *our* relation to *the truth* is in focus, e.g. because *stating* the truth is contrasted with *concealing* it, or *knowing* the truth with *ignoring* it, the determinate article is added, or the noun *alētheia* (with article) is used: *ta alēthē/tēn alētheian legein* (e.g. *Apol.* 18a6, *Gorg.* 487a6).⁷ Yet note also that this talk of “the truth” is elliptical. If the context does not indicate what the speaker has in mind, an additional specification is required: “the truth regarding/about (*peri*) . . .” The locution “the truth” can also convey the idea of *relevance*. Not every correct piece of information with regard to an object or situation contributes equally to an account of “the truth (about it)”.

Accounts of certain situations, events, etc. can be called more true or less true on account of how accurately they represent the facts. I think we should acknowledge some analogy between how such accounts can be *more or less* true to the facts and how depictions can represent objects or situations more or less accurately. One might speak in such cases of *representational* truth or accuracy. How exactly this relates to the basic notion of propositional truth is open for debate.

1.1.2 Epistemic truth

By “epistemic truth” I mean truth as that which can come to be known in the acquisition of knowledge. The use of truth vocabulary in connection with verbs of knowing

as in “It is as you say,” while the existential usages have different kinds of derivation not involving comparison (Kahn 1973/2003: 228–370, esp. 342–55). However, we should distinguish between two forms of veridical *einai*: one that is exemplified by the phrases “It is so” and “It is as you say” (acceptable also in English), and another that allows for the “is” (*esti*) to be predicated of the content of an assertion or belief *without a comparative clause*, e.g. “What you say, is (*i.e.* the case)” — which I call the veridical “is” *proper* (uncommon in English). The veridical participle *onta* in object position clearly is derived from this latter type. Kahn postulates, but does not prove, that the latter usage is only an elliptical variant of the one with a comparative phrase, and that for this reason it is clearly distinct from existential usages. I don’t think his argument succeeds.

⁶ Although in general the *proper* veridical usage of *einai* (“to be”) in Greek is hard to reproduce in English, we may compare it to the use of the future form “what will be” in the refrain “whatever will be, will be.” This phrase can cover the whole range just mentioned. Its Greek counterpart is the future participle *ta esomena* (as in the Homeric formula for what a seer can predict, e.g. *Iliad* 1.69–70).

⁷ On the pragmatic notion of “focus” compare, for instance, Kadmon (2001: 227ff).

and inquiring seems, at first sight, a simple extension of its use with verbs of saying and believing (e.g. *Rep.* 581e1, *Crat.* 484b6). Yet since a verb of knowing usually entails that the person has a true belief, it does not need to be stressed that the knowledge in question is, in fact, true or knowledge of something true. What the statement focuses on is, rather, whether *the* truth in question (use of the definite article!), i.e. the truth about a certain person, a certain situation, etc., is known rather than ignored, or known fully rather than partially, etc.

As with verbs of saying and believing, there is, I think, a noticeable tendency in Greek thought to identify the truths to be known with those aspects of reality which can be represented in our knowledge-like cognitions (i.e. certain states of affairs, patterns of change, objects such as they really are, etc.). When epistemic truth is clearly identified with an aspect or element of reality, I will call it the *ontologico-epistemic* notion of truth.

1.1.3 Adverbial and attributive truth

Propositional clauses can be supplemented with an adverbial truth expression (“truly” or, in English much more frequently, “really”). In its basic form, this usage serves to put some emphasis on the assertion, question, hypothetical clause, etc. without modifying the propositional content. Yet there is also a usage that has the effect of, as we might say, setting the applicability bar higher, such that we would exclude defective or less perfect examples from the application range of the predicate in question. This latter usage is *not* descriptively neutral. We can distinguish these two cases as “confirmative emphasis” and “paradigmatic emphasis.”

The adverbial truth expression is obviously derived from propositional truth since its most basic form simply stresses the truth of an ascription. From the adverbial form, it is only a small step to the attributive usage of the adjective *alēthēs* (and its cognate *alēthinos*).⁸ Some x is a true (real, genuine, authentic) F if it is *really* the case that x is F. This may, again, simply emphasize the truth of the ascription without modification of the descriptive content, but it may also have the force of paradigmatic emphasis.

1.1.4 “Truth” or “reality”? “Ontological truth”?

There is no distinct word in Greek that would correspond to our expression “reality” (derived from a neologism in medieval philosophy). The usual Greek expression to serve this purpose is the participle of “to be” (*ta onta*, “the things that are”), which is used, as I have pointed out, without a sharp boundary between the veridical and existential

⁸ On the use of *alēthinos* compared to *alēthēs* in Plato, see Szaif (1996: 50 fn. 43).

connotations. We have also noted that sometimes “reality” and its cognates, rather than “truth” and its cognates, are the most fitting translations for “*alētheia*” and its cognates. In fact, we can observe a far-ranging parallelism between the words of being and truth in Greek such that there is a significant number of instances in which both the vocabulary of being and that of truth bear a close relation to our notion of reality—including cases in which the truth of an assertion or belief is in focus.⁹ Does this indicate a genuine difference between the Greek notion of *alētheia* and “truth” in English? If we had to locate the quality *truth*, or also *the truth* understood *concretely* (i.e. as something that *has* truth), either on the word-side or the world-side of the word/world divide, I suspect we would locate it on the word-side, for instance as a property of the *content* of certain beliefs or assertions. In Greek usage, on the other hand, it seems perfectly acceptable to locate *the truth* described or known on the world-side so as to identify it with some aspect or part of reality. I submit that to this (limited) extent there actually is a difference between the ancient Greek and the modern understanding of the truth terminology—a difference which has had some influence on the development of theories of propositional truth, for example when Aristotle in some texts seems to suggest that something like states of affairs are the primitive truth-bearers.

Let me add, at this point, a brief comment on the term “ontological truth,” which is used in discussions of the history of the philosophical concept of truth, frequently as an antonym to “logical truth.” This terminology is also based on the world/world distinction such that cases in which truth is attributed to items on the world-side are classified as instances of ontological truth. I’ll be using the expression “ontological truth” occasionally, but note that this expression can mislead us if we forget that there are very different ways in which a truth-ascription can relate to something on the world-side—for instance when certain *objects* are called true in light of their genuineness and non-deceptive quality, but also when states of affairs are viewed as the primitive propositionally structured truth-bearers.

1.1.5 Moral truth

The adjective *alēthēs* can also connote *truthfulness* as a moral character trait. I won’t comment here on the role of the moral meaning of *alēthēs* in philosophical virtue theory. Yet note that this connotation can be present as a metaphor when objects are called “true” in contrast to “deceitful” objects that are a source of misleading appearances.

⁹ In the case of propositional truth, (*ta*) *onta* can be used instead of (*ta*) *alēthē* (or *tēn alētheian*) in the object position after a verb of saying, believing, or knowing. In the adverbial case, either *ontōs* (an adverb derived from *einai*, “to be”) or (*hōs*) *alēthōs* (an adverb derived from *alēthēs*) can be used, often corresponding to the word “really” in English. Finally, in connection with the knowledge vocabulary, *alētheia* can be used to denote the reality behind the appearances—the true reality, as it were, which the quest for knowledge wants to reveal. See also Vlastos (1965: 59f); Kahn (1978: 35f).

1.2 PLATO'S CONCEPTION OF TRUTH IN THE CONTEXT OF HIS THEORY OF FORMS

Beginning with his early dialogues, Plato consistently maintained that philosophy, as a practical attitude, is an unwavering commitment to the truth.¹⁰ In his middle-period dialogues this theme becomes linked to what I call his ontologico-epistemic concept of truth (cf. section 1.1.2), which is a central element in his metaphysical epistemology of knowledge as “knowledge of Forms.” The most important body of evidence can be found in the central books of the *Republic* (V–VII), but other dialogues, including the *Phaedo*, the *Symposium*, *Republic* X, the *Phaedrus*, the *Timaeus*, and the *Philebus*, contribute to our understanding of this Platonic notion. In the central books of the *Republic*, the first step in the argumentation concerning the nature of philosophy culminates in the programmatic assertion that the philosopher is a “lover of the sight of truth” (*philothēamōn tēs alētheias*, 475e4). The subsequent arguments and similes frequently involve the characterization of the “Forms” as that which is true and as the realm of truth. The “simile of the sun” (506d–509c) provides a particularly strong piece of evidence for the central role of this notion, while the immediately following similes (Divided Line, Cave) confirm it and, together with other arguments in the whole sequence, provide more detail on how this notion relates to Plato’s views on scientific methodology, education, and rational governance. There is no space here to discuss Plato’s use of truth terminology in the context of individual dialogues (although this would be preferable). Instead I am going to present a systematic overview of the results. Since Plato’s approach to the notion of truth does not come out of nowhere, I will begin with some remarks on the developments that took place among his philosophical predecessors.

1.2.1 The views of Plato’s predecessors on truth

In early Greek philosophy, we can identify two interrelated ideas that lead to an interest in the concept of truth. First, beginning with Xenophanes (21B34), there is the theme of how difficult, if not impossible, it is to achieve a precise and certain grasp of the truth owing to the weakness of human cognitive powers. In this connection, Xenophanes (ibid.) uses the formulation that unreliable “appearance” is “spread over everything,”¹¹

¹⁰ Cf. *Gorgias* 526d (see also 525a3, 526c1–2). For philosophical *love of truth* as the foundation of virtue in general, cf. *Rep.* 485a–487a, 489e–490d, *Legg.* 730c. The entire educational program for the philosophers in the *Republic* is dedicated to the ascent to the truth (e.g. 490a–b, 519b, 525b1, 527b9). Couched in a mythical description of the souls’ struggle for a vision of the true exemplars (247c ff) and for nourishment from this “plain of truth” (248b6), this theme also informs the *palinode* speech in the *Phaedrus*.

¹¹ Reading *pasi* in B34, line 4, as a neuter.

which already points to the second idea, *viz.* the opposition between appearance and some underlying truth or, more generally, the true nature of reality “behind the appearances.” This contrast can be spelled out epistemologically or ontologically. Epistemologically, convictions that are true and based on sound reasoning are set in opposition to inaccurate or misleading beliefs. Ontologically, true reality, which is to be revealed through some superior reasoning, is contrasted with the world of appearances—roughly the world as it appears to us and is represented in our common ways of thinking and speaking (e.g. Heraclitus, 22B1, B2, B17, B34, B72, B114; Parmenides 28B1.28–30, B6, B8.38–41, 50–2, B19).

In connection with this theme we find clear evidence of the influential idea that “the truth” is something that transcends our individual or species-bound perspectives and is also not restricted by temporal qualifications. In Parmenides’ philosophical poem, the allegorical introductory part ends with a goddess’s promise that the listener will, first, learn the “unshaken heart of persuasive/trustworthy *truth*” (B1.29),¹² but subsequently also the “appearances among mortals” which “do not contain any *truthful* conviction” (or “*genuine* credibility,” B1.30; see also B8.50–2). When the first account is said to be “about (*amphis*) the truth” (literally: “around the truth,” “attached to it from both sides” B8.51), we should note that the word “truth” (*alētheiē*) is very likely used in the sense of “true reality” and identified with being. We find a kindred usage of truth vocabulary in a passage that states about the “mortals” that they attributed certain features to reality because they took these features (generation, locomotion, etc.) to be “real/true” (*alēthē*) (B8.39). According to Parmenides, the only *basic* feature that they *should* have accepted as real/true is *to be*. Since Parmenides also rejects temporality with respect to *that which is* (at least according to the common ancient understanding of time, which links time to change and motion), we are getting a glimpse of the influential idea that truth excludes temporality¹³ and that temporal facts, therefore, fall outside the scope of truth strictly speaking (B8.5f, 19f).¹⁴

Passages like these point to what I am calling the ontologico-epistemic concept of truth, which will become very prominent in Plato. According to this usage, the word “truth” (*alētheiē*) does not refer to a quality of statements or mental states but to *true reality*, which can only be revealed through reasoning and is opposed to the deceitful appearances in how we perceive the world. This is an *epistemic* notion of truth because of the conceptual link between truth (*concrete sense*; cf. section 1.1.4) and knowledge; and it is also an *ontological* notion since it identifies *the truth* with knowable *reality*. Another feature anticipating Plato is that this Parmenidean notion incorporates a set of ontological criteria that a presumed case of truth (or knowable reality) has to fulfill—criteria such

¹² Text with “*eupeitheos*” following Sextus Empiricus *e.a.*

¹³ His talk of the “*unshaken* heart” of truth in B1.29 also alludes to the view that truth, or true being, cannot be subject to destruction, change, or motion.

¹⁴ See also Melissus, 30B8: the Eleatic author seems to presuppose in this argument that a *presumed* element or feature of the world cannot be an instance of knowable truth, or a true element of reality, if it is *temporary*.

as indestructibility, exemption from any kind of change, and the exclusion of opposite characterizations.¹⁵

This emerging metaphysical notion of truth or true reality is countered, on the one hand, by various relativistic and nihilistic arguments about truth invented by some of the major sophists (beginning with Protagoras and Gorgias) and, on the other, by the theories of the physicists who try to save the phenomena of nature while incorporating some of the basic Parmenidean (and Zenonian) arguments. In general, we can understand the positions of Parmenides and of subsequent sophists and physicists as different answers to the question of whether there is some objective truth, or reality, for us to know and, if so, how it relates to appearances: Parmenides drew a sharp boundary between *the truth* (true reality) and the phenomenal world, reducing the latter on metaphysical grounds to the status of deceitful appearance (despite the fact that he also thought that an internally plausible scientific account of the structure of the phenomenal world was possible). Protagoras countered this with an account that *identified* truth and appearance by relativizing truth, or reality, to the perceiving person.¹⁶ In Gorgias' treatise *On Not-Being* (and also in Xenias)¹⁷ we find arguments, entertained at least dialectically, that amount to a nihilistic position regarding truth/true reality, yet without a Protagorean endorsement of appearances as relativistically true. In response to these detractors of objective truth, the atomist Democritus (whose fragments provide the most interesting material on truth among the post-Parmenidean physicists) asserted the unreality of basic perceptual objects (cf. 68B6, B9, B117 [=D19, 16, 15 Taylor]), yet postulated the existence of some underlying true reality that can be revealed through reasoning (cf. 68B11 [=D22 Taylor]) and that provides the explanatory basis for the phenomena observable in nature. Like Parmenides, he attached certain ontological characteristics to basic entities that qualify them as true reality, such as

¹⁵ The ontologico-epistemic usage of truth terminology is not the only one in Parmenides. Other usages are exemplified by the expression *alēthēs pistis* (B1.30, B8.28: "true conviction" or "true/genuine credibility") and by the complex metaphor in which the path of inquiry itself is called true (B8.17f, *alēthēs, etētymos*; cf. B2.4).

¹⁶ He gave a brief theoretical formulation of his relativistic notion of truth in a work that bore the title "Truth" and began with a statement of his "Man is the Measure" doctrine: "Man (the/a human being) is the measure of all things, of those that are that they are, of those that are not that they are not." (The expression "those that are" is best interpreted in a vague sense that comprises existence and veridical being, perhaps even copulative being [x is F]). As a challenge to our common assumptions, this statement suggests that it is *not the things that are* which function as the measuring-rod for the truth and accuracy of our thoughts; we ourselves, rather, and the ways in which we perceive reality are the measuring-rod for what is to count as real or true. Truth is, hence, relative to the perceiver. This relativity can be interpreted either as species relativity or as individual relativity. Yet the individual-relativity reading agrees better with what seems to have been one of Protagoras' own examples (cf. Plato, *Tht.* 152b–c): the different perceptions of the same wind as cold or not cold depending on who the perceiver is. In the same context, Protagoras also seems to have stated that however things *appear* to us, they *are* (or are true) for us (*Tht.* 152a, 169e–170a [cf. 171b6f]; cf. Lee 2005: 12f, on these passages as evidence for the historical Protagoras). This claim suggests a broad application of his relativistic principle to all modes of epistemic appearance, not just sense-perception.

¹⁷ Cf. Sextus Empiricus, *Adv. Math.* II, 53f, 399.

indestructibility and absence of qualitative change. Against Parmenides, however, and similarly to other physicists, he connected being and appearance by way of a physicalist account of thought and perception.¹⁸

Plato's position is, in at least one important respect, similar to the atomist approach. Like Democritus he makes use of the Parmenidean (and Heraclitean) contrast between true reality and appearance in order to establish a domain of truth behind appearances and present it as the proper subject-matter of scientific knowledge, accessible only through reasoning, yet providing an explanatory foundation for the phenomenal world. He introduces, however, a radically new type of entity to fulfill the grounding function—no longer material atoms, but incorporeal entities that function as graspable contents and in which physical objects “participate” without being reducible to them.

Another important influence on Plato's understanding of knowledge and truth comes from the development of Greek scientific mathematics. As Plato points out early on, it is through counting and measurement that disputes become decidable, because quantification provides a basis for secure and indisputable judgment (*Euphr.* 7b–d, cf. *Prot.* 356c–357a, *Phil.* 55e–56c). The practice of mathematical proofs, moreover, nourished the expectation that also in philosophical matters it would be possible to produce exact and secure results, and rational agreement, through sustained abstract reasoning. Moreover, since Greek mathematics (arithmetic and geometry) involves the idealization of natural numbers and geometrical figures, it provided an additional source of inspiration for the idea that the truth (true reality) to be revealed by rational investigation was something beyond perceptual appearances, a realm of pure and everlasting objects of thought.¹⁹

1.2.2 The use of truth terminology in Plato's theory of Forms

When Plato wants to highlight that the Forms are the only objects that allow philosophers to fulfill their quest for knowledge, he calls the Forms *that which is true*, or (the realm of) *the truth*, while empirical objects are presented as deficient with respect to

¹⁸ 68B9, B11 (=D16, D17, D22 Taylor) together with testimonia 178a, c, 179a, d Taylor.

¹⁹ Plato's emphasis on exactness (*akribēia*) as a mark of truth or knowledge of the truth (e.g. *Rep.* 504b–c, d–e, 529c–d (cf. 530a–b), *Phdr.* 261e6–262b4, *Tim.* 29c3–d7, 52c6, *Polit.* 284c–d, *Phil.* 57b5–d2, 57e3–58a6, 58b9–c4, 58e4–59b9) is probably rooted in his experience of mathematical study. He also likes to use mathematical definitions and proofs as examples of a scientific approach (e.g. *Men.* 75b–76a, 82b–85b, 86e–87b, *Phd.* 92c–e, *Thet.* 147c–148c, 162e). In the *Republic*, he criticizes mathematics for its inability to understand and prove its own conceptual foundations (and perhaps also for its partial reliance on visual diagrams) (509d–511e, 533b–c), yet also insists on its function as the indispensable gateway to philosophical science (“dialectic”), giving a great deal of space to mathematical study in his philosophical curriculum not least because it “turns the soul away” from perceptible objects to the *realm of truth* populated by objects of pure thought (518b–519b, 524d–532c, 537b–c; see section 1.2.2).

truth. This conception has often been dubbed an “ontological notion of truth.” The task for us is to analyze what “true” and “truth” mean in this context.²⁰

Plato’s metaphysical epistemology sets out with a claim concerning conceptual understanding according to which our words and concepts, or at least a subset of our words and concepts, relate to objective essences. To put it in modern terminology: for any given general term “F,” the philosophical inquiry will begin with the heuristic assumption that “F” corresponds to an objective essence (Fness or *the F-itself*) and that full understanding of the term “F” requires that one reach a correct definition of this essence. Plato calls such an essence a “Form” (*eidōs, idea*).²¹ The F-Form, as the pure reified content designated by the term “F,” is said to *be F* in its own right, itself by itself, while the many things that are commonly called F (e.g. just, or triangular) share in this kind of being only derivatively—not by themselves but in virtue of an ontological relation that Plato labels as “participation in a Form.” These semantic and metaphysical claims go hand in hand with a theory of *scientific knowledge* (*epistēmē*) that identifies the realm of Forms as the exclusive object range of *epistēmē* proper, at least according to the traditional reading that I am endorsing. (Note that the scientific knowledge in question is meant to relate to eternal and universal truths. In other contexts Plato is willing to use a broader notion of knowing that allows for knowledge concerning individual empirical objects and facts.)

Since Plato’s notion of truth is closely connected to his concept of *epistēmē*, I have to comment on the question of the propositional character of Platonic *epistēmē*. First, it is not possible clearly to distinguish between the object and the content of knowledge in the case of knowing a Form, because the Forms are nothing but reified essences, hence reified descriptive content, and Platonic *epistēmē* is the attempt to acquire clarified and objectified concepts in accordance with the Forms. Secondly, conceptual understanding is, by definition, not propositional knowledge, since concepts serve as the primitive building blocks for our propositional beliefs. Yet Plato also maintains, at least as a heuristic principle, that a Form should be viewed as definable, and while the definiens is not a proposition, the definition which links definiendum and definiens is. Thirdly, definitions are not isolated but interconnected in complex taxonomies which philosophical dialecticians have to investigate. Thus the knowledge Plato’s philosophers are striving for is both an enhancement of our pre-propositional *conceptual grasps* and a complex

²⁰ Good examples for the application of the notion of “that which is true” to the objects of intellectual grasp or investigation, and especially to the Forms, can be found at *Symp.* 212a5, *Phd.* 67b1 (cf. 66d7), 84a8, *Rep.* 484c9, 519b4, 529d1–3, 539c6f, *Phdr.* 247d4, 248c3f, 249d5. Examples for referring to a Form, or the realm of Forms as a whole, as “(the) truth” (whereas the sensibles are dubbed, for instance, “the [realm of] becoming”), include: *Rep.* 501d2, 525b1, c6, 526b2f, 527b9, e3, 585c2, 597e7 (cf. 602c2, 603a11, 605c3), *Crat.* 439a4 (in reference to the designatum of a general term), *Phdr.* 247c5f (cf. c6–8), 248b6, 249b6. Compare also the use of *alēthēs* and *alētheia* in the similes of the Sun (*Rep.* 508d4–509a7, cf. 475e4, 517c4), the Divided Line (*Rep.* 511e, cf. 510a9), and the Cave (*Rep.* 516a3, cf. 515c2, d6f). Instead of *alēthēs*, Plato he sometimes also uses the cognate *alēthinos*, (compare e.g. 529d2 and d3).

²¹ I am going to follow the convention of using the word “Form” with a capital letter to designate a Platonic *eidōs*.

theory of the network of Forms that would have to be expressed propositionally.²² These data create some ambiguity regarding the propositionality of Platonic *epistēmē*, which also affects the corresponding notion of *alētheia* qua reality *to be known*.²³

A relevant linguistic phenomenon, in this connection, is the interchangeability between the adjective *alēthes* used as a substantive (*to alēthes*, *ta alēthē*: “that which is true,” literally “the true”) and the abstract noun “truth” (*alētheia*) used *concretely* (cf. section 1.1.4). We may compare this with a general grammatical feature in Greek thanks to which a substantival adjective and the corresponding abstract noun can be used synonymously (e.g. “the hot” and “heat”). Accordingly, Plato calls *that which can be known* (the reified content of *epistēmē*) both “the true” (*to alēthes*, *ta alēthē*) and “the truth,” without any significant difference in meaning. The default would be to interpret such cases as instances of the *epistemic* usage of truth vocabulary (cf. section 1.1.2). Yet there is an additional complication. A closer look at some key passages reveals that they often involve an ambiguity between the *epistemic* and the *attributive* usages (cf. section 1.1.3).²⁴ In addition, a quasi-moral sense of “true” (cf. section 1.1.5) can also be in play: *that which does not deceive*. I submit that Plato’s use of truth vocabulary in such passages is meant to encompass all of these connotations since in his view the *reality to be known* (ontologico-epistemic truth) is made up of objects which exhibit attributive truth (e.g. in virtue of their purity or in virtue of being the original exemplars) and which, moreover, thanks to their attributive truth manifest themselves in a non-deceptive manner (dependability, quasi-moral truth). In a modern discussion, we would distinguish sharply between an object *o* that is a true *F* and the true propositional content “*o* is *F*.” Only in the latter case would we be willing to speak of a truth that can be known. In the key passages for Plato’s ontologico-epistemic notion of truth, however, these distinctions become blurred because of the factors mentioned in this and the preceding paragraph.

With respect to attributive truth (genuineness), we can distinguish two types of usage that are in play when Forms are called *that which is true*. The first of these relates to the distinction between “originals” and mere copies or semblances (*eidōla*). Plato rather often contrasts “the true” (the original, the real thing) with its semblances (*eidōla*),

²² For an instructive modern discussion of the implications for Plato’s concept of knowledge cf. Burnyeat (1980); Fine (1979); Nehamas (1982). The contribution by Nehamas is particularly forceful in explaining why Plato’s concept of *epistēmē*, with its focus on knowing essences, cannot be reduced to a general notion of propositional knowledge (cf. Fine 1979 for a different view). Note that this approach does not have to entail a commitment to the notion of knowledge by direct acquaintance. There are good reasons for believing that Plato’s notion of *epistēmē* is shaped by his concern for the grasp of essences, yet since Plato takes argumentative enquiry (“dialectic”) to be the sole vehicle for the advancement of our understanding of essences, this kind of knowledge has to be grounded in discursive reasoning.

²³ Cf. *Parm.* 134a–b, a passage that assumes that each species of knowledge correlates with a Form (knowledge of justice, knowledge of magnitude, etc.), while the universal genus *knowledge* has truth as such as its defining object. This implies that the Forms are understood as specifications of “the truth.”

²⁴ These two aspects, the ontologico-epistemic sense and the attributive sense, correspond roughly to the two senses of ontological truth discussed by Hestir (2004) (compare now also Hestir 2016: 234–8).

and in quite a few cases this is explicitly applied to the Forms qua exemplars in contradistinction to their imperfect and transient instantiations in the sensible realm.²⁵ In this connection, he also describes the ontological relation of participation as a relation between an original and its copies or semblances.²⁶ This includes the idea that empirical instances, qua semblances, are *untrue* insofar as they conceal the fact that they are only copies (*Rep.* 475e–476d). As long as we are unable to grasp the Form as the underlying formal cause, we are under the sway of the appearances and take the bona fide empirical instances of Fness to be the only (or primary) kind of reality denoted by the term “F.”

The second kind of usage contrasts items that are unmixed and pure with things that are subject to mixture, especially mixture of opposites. When Plato wants to mark the difference between Forms and their empirical instances, he often points out that an empirical instance of the Form F-itself (i.e. an object that partakes in the Form) can also manifest itself as the opposite of F (i.e. its polar contrary), say G, or at least as not-F. This mixture of conflicting appearances is also the reason why empirical entities (tokens or types) are said to lack clarity (*saphēs, enargēs*) and to be obscure (*skotōdēs*).²⁷ Conflicting appearances result from changes in the observer’s perspective, focus, or (implicit) comparison class. To paraphrase some of the Platonic examples: a perceptible object looks different from different angles (*Rep.* 598a), say *round* from one angle, *rectangular* from another;²⁸ it may look *equal* to some other object if seen from further away, *unequal* if seen close up;²⁹ *large* if compared to grass, *small* if compared to mountains;³⁰ *cold* if you are heated, *warm* if you are cold;³¹ *one* qua member in a set of objects

²⁵ e.g. *Symp.* 212a2–7, *Rep.* 520c1–d1 (cf. 484c6–d3), 597a1–b1; see also *Rep.* 510a8–10 (in the simile of the Line) together with 511e, and *Rep.* 597d11–598c4 (esp. 598b1–8), 599d2f, 600e4–601b1, 605a8–c4. For the broader contrast between *to alēthes* (“that which is true/real”) or “the truth” (*alētheia*) and words such as “image,” “semblance,” “shadow,” or “imitation,” see *Men.* 100a2–7, *Lys.* 219c5–d5, *Rep.* 382b7–c1, 533a1–3, 583b3–5, 586b7–c5, 587b14–e4, *Crat.* 439a5–b3, *Tht.* 150a8–c3 (cf. e6f, 151c2–d3), *Soph.* 234c2–e2, 240a4–b2, *Pol.* 300c5–7, *Legg.* 889c–d.

²⁶ Plato’s practice of calling the F-Form the F-itself (e.g. “the just itself”) already points in this direction. Compare also how Plato uses the phrase “that (itself) which IS F” (i.e. emphatically *is F*) as the general label for a Form in contradistinction to its empirical instances (e.g. *Phd.* 75a11–b2, c10–d3, *Rep.* 597c1–d3, *Parm.* 133d7–134b1) and connects this with the model/copy analogy: *Rep.* 597a1–5, *Crat.* 389a5–b8, d4–8.

²⁷ For clarity (*saphêneia*) attributed to the Forms, or objects of thought, cf. *Rep.* 509d9, 515e4; see also *Phd.* 83c8, *Rep.* 511a7 (cf. 484c7) (with *enargēs*). (*Saphêneia* is also, and perhaps more typically, an attribute of knowledge in relation to such inherently clear and precise subject matters; cf. *Rep.* 511e2–4, *Phil.* 57b–c.) On the cognitive “darkness” (obscurity) of sensible objects compare *Rep.* 479c6–d1 and the central similes of the *Republic*, starting with the Sun (s. a. 518c7f, 520c4). This talk of clarity versus obscurity is sometimes also associated with the degree of *exactness/accuracy* in how a certain determination is realized in an object (cf. passages cited in fn. 19). Lack of exactness is one source of conflicting appearances.

²⁸ See also *Soph.* 235e–236a on the effects of different angles of visual perception.

²⁹ Conflicting appearances as a result of changing distances: e.g. *Prot.* 356c–e, *Rep.* 602c7f.

³⁰ The effect of different (implicit) comparison classes on judgments: e.g. *Hipp. ma.* 287e–289c (using the example of beauty), *Rep.* 523c–524a (large or small, hard or soft, etc.), *Phd.* 102b (large or small).

³¹ This harks back to the classic Protagorean example mentioned in *Tht.* 152b. The list of relativizing respects given in *Symp.* 211a2–5 also mentions observer relativity (*tisi men . . . , tisi de . . .*).

of the same kind, *many* on account of its parts (cf. *Parm.* 129c–d, *Rep.* 525a, d–526a); and a type of action may turn out to be just in some situational contexts, unjust in others (e.g. *Rep.* 331e–332a).³² Yet the Form or reified content as such, for instance largeness or the *large itself*, is always clearly separated from its opposite, e.g. the *small itself* (cf. *Rep.* 523a–524d [cf. 605b–c], *Soph.* 252d). The conceptual link between the Forms’ unmixed being and the truth terminology remains implicit most of the time. Yet in *Phaedo* 65a–67b Plato explicitly equates the claim that Forms are pure/unmixed with the claim that they are something true (67b1). The underlying general idea is that the purest specimen is also the truest, and that the Form represents a notional content in its purest form, not obscured by other features that don’t pertain to it (see also *Symp.* 211a–b, d–e). The *Philebus*, too, elaborates on the conceptual link between purity and attributive truth (52d–53c, 58c–d, 59c).³³

The instability of appearance in the case of empirical instances is a function of changing perspectives and does not require real change in the object. Yet Plato probably also views permanence as an aspect of *genuine* determination, in the wake of the idea already prevalent among the Eleatics that genuine truth is incompatible with change in the object; for real change too forces us to relativize our judgments, *viz.* to time.³⁴ I suspect that Plato took beliefs about processes and changeable attributes to involve the distinction between past, present, and future (cf. *Tim.* 37e–38a), which would mean that temporal qualification also introduces a subjective component in that it presupposes the temporal standpoint of the enquirer.³⁵

Plato’s comparative ontology of ideal Forms and empirical instances includes what has been dubbed a theory of “degrees of reality.”³⁶ In various contexts, Plato attributes

³² This list as a whole recalls the relativistic (or subjectivist) account of the truth of perceptual judgments in Plato’s reconstruction of Protagoreanism in the *Theaetetus* (the “secret doctrine”). Plato can hardly have endorsed the extreme form of relativism set out there for all types of judgments about individual empirical objects, since that would have thwarted his ethical agenda. If, or to what extent, he endorsed it for a more narrowly defined class of perceptual judgments (cf. *Th.* 179c) is a contentious issue.

³³ Plato sometimes seems to conflate, or blend, the two criteria of being impure and being a mere copy or semblance, e.g. in *Rep.* 583b3–5 (cf. 584b–c, 586b–c) and 382b7–c1.

³⁴ The list in *Symp.* 211a2–5 of relativizing respects that result in conflicting appearances mentions relativization to specific times alongside such aspects as what the object is being compared to or where it is located. (Irwin 1977, by contrast, maintains that Plato’s flux thesis does not include reference to real change.)

³⁵ Note that without reference to temporality and real change it would be much more difficult to exclude instances of kind terms like “human being” or “couch” from the realm of ontologico-epistemic truth, since such ascriptions are not context-dependent in the same way as “large” or “beautiful” are.

³⁶ On degrees of reality, cf. Vlastos (1965, 1966) and, on the semantics of “being” involved here, Brown (1994). In broad outline, I agree with Vlastos’s interpretation of the higher reality of Forms as a function of certain ontological qualities that result in higher cognitive dependability. (Vlastos adds another aspect which he calls evaluative.) My interpretation tries to give a more thorough account of the various semantic aspects of truth terminology that underlie Plato’s talk of the being and truth of Forms. Code (1993) claims that Vlastos’s account produces an inconsistent theory, since the two criteria of purity and essential being (the latter corresponding roughly to what I have called underived being) are inconsistent in that an item has to be, at least in part, essentially F in order to be impurely F (ibid. 91–3). Yet (Vlastos’

being exclusively to the Forms, as he does with ontological truth.³⁷ Yet since he does not mean to suggest that perceptible objects are mere figments, he also uses expressions that entail *degrees* of ontological truth and parallel *degrees* of being.³⁸ In this context, *being* is to be understood, roughly, as *being something determinate* (being F), and the higher degree of being (and ontological truth) on the part of a Form reflects the fact that the Form exhibits its defining content in a more genuine and reliable way than the corresponding perceptible objects, which also makes it a more fitting object for the enquiring mind.

My comments in this section so far should make it easier to understand the ways in which truth is presented as an attribute of cognitive objects in the *Republic's* famous simile of the Sun (506d–509c). A close analysis of this simile (and its connections with *Rep.* 476eff) is not possible here. Suffice it to say that the simile compares the effect of *truth and being* as attributes of an object of philosophical inquiry to the effect of sunlight in the realm of visible objects. It suggests that an object of inquiry requires ontological truth to become knowable, just as colors require sunlight (as opposed to dim moon light) to become clearly and reliably visible. The guiding idea, again, seems to be that the lack of ontological truth characterizing empirical instances results in an inferior state of cognitive representation in our mind, since we cannot yet latch on to some clear and reliable content (i.e. some content not obscured by conflicting appearances). The Form, on the other hand, being this content itself, purely and underived (attributive truth), is cognitively dependable thanks to this ontological quality and can thus function as the subject-matter of scientific knowledge (i.e. serve as the domain of truth in the ontologico-epistemic sense).

The simile also claims that truth and being originate from the Form of goodness. The role of the Good, as the root of truth and being, is susceptible to various strategies of interpretation.³⁹ Yet given what we have said about the role of the concepts of truth and being in this simile, the safest approach would be to suggest that one of the functions of the Form of the Good is to represent, in the most general way, a state of *perfection*. Since the Forms are called true in light of the fact that they are what they are in an undiluted, faultless, non-derivative, unchanging, and, hence, reliable way, such

exegesis aside), what Plato means to convey is that an empirical object that is F only by participation (i.e. derivatively) is, for the very same reason, also susceptible to the co-presence of the opposite of F (at least in cases where there is an opposite) and, hence, to an impure mode of being-F.

³⁷ The only difference, as a matter of fact, between calling Forms things that *are*, and calling them *that which is true*, is that the latter expression is better suited to highlight the contrast with deceiving appearance, whereas the former is often used in opposition to alteration and “becoming.” Yet the contrast with appearance is also at play in the terminology of being, while the contrast with alteration and becoming is also relevant for Plato’s notion of the Forms as *the true*. (On the connections between the terminologies of being and truth in Plato compare also Kahn 1981.)

³⁸ Cf. *Rep.* 515c–d (compare d2–4 and d6–7), 479c6–d1, 511e, 585b–d; *Tim.* 51b6–d7.

³⁹ On Plato’s conception of “the good” compare for instance Burnyeat (2000) (with whose reconstruction of the Platonic conception of intrinsic value I mostly agree; cf. Szaif (1996: 278–87) and Santas (2001: 167–93).

truth is a kind of ontological perfection, and qua perfection it requires the Good as its formal cause.⁴⁰

Finally, it should be mentioned that Plato's metaphysical epistemology can refer to truth also as an intellectual state ("truth in the soul"). This notion complements the ontological truth conception just described and cannot be reduced to true belief.⁴¹ For in this specific usage, "truth" is juxtaposed with the notion of intellectual understanding (*nous*, *phronēsis*) and apparently almost identical with it (cf. *Phil.* 65d). It characterizes a cognitive state in which the soul is no longer attached to empirical semblances but has achieved full cognitive clarity with respect to the proper objects of knowing.

1.2.3 Does Plato distinguish degrees of propositional truth?

As mentioned, Forms in Plato have an explanatory role with respect to empirical objects in that the latter exhibit their transient determinations in virtue of "participating" (having a share) in the Forms. Closely linked to this explanatory role of the Forms is their function as standards or measures to be applied in our judgments about empirical objects. Qua *exemplars*, they are the truth against which we measure if, or how well, a particular instance exemplifies a Form.⁴² These observations about the role of Forms as standards or criteria to be applied in our judgments take us to the topic of truth as a property of judgments (beliefs, statements). Notwithstanding the fact that Plato's ontologico-epistemic notion of truth conceives truth as a feature of the realm of Forms, he never rejects the common usage of truth terminology in connection with expressions of saying and believing. But what implications, if any, does restricting unqualified

⁴⁰ The simile of the Sun would also provide the best evidence for the claim that Plato's notion of truth is linked to an etymological interpretation of the word *alētheia* as "unconcealedness"—a claim that used to exert some influence among European scholars. The comparison of ontological truth with light rendering colors clearly visible can indeed suggest the idea of "unconcealedness." I also think it is possible, and even probable, that *in the context of this simile* Plato intended to evoke such an etymological association to enhance the illustrative effect of the simile. Plato is willing to use etymological associations for illustrative purposes in other contexts as well. Yet this would not vindicate the much more radical claim that his truth-conception in general is somehow determined by an etymological association (cf. Szaif 1996: 132–52).

⁴¹ The most telling passage that evidences this meaning is *Rep.* 490a–b; see also *Phd.* 66a, *Tim.* 29c (*alētheia* substituting for a term like *epistēmē*, cf. *Rep.* 533e–4a), *Phil.* 65d. (*Men.* 86b1 requires a different explanation.) There is also a virtue-ethical meaning of "*alētheia* in the soul," which I am not discussing here (cf. Szaif 2004).

⁴² Cf. *Rep.* 520c on applying one's knowledge of the ideal truth in one's judgments about empirical instances or "semblances"; *Phd.* 74d–75b on how empirical objects can fall short of the Form they exemplify; *Rep.* 473a, 484cd (cf. 500d–501c, 540a–b) on how the Form (e.g. of justice) functions as an ideal truth to be emulated in our practice and our political institutions, which will, however, always fall short of this truth. See also *Phil.* 62b on the notion of a "false circle"—false in virtue of being inaccurate, while a fully accurate ("true") circle exists only as an object of thought. (Against Nehamas 1975 and others, I would maintain that the inaccuracy of empirical instances and the idea of approximation to the Form as an ideal standard is *a*—but not the only—relevant aspect in Plato's comparative ontology.)

truth (in the ontologico-epistemic sense) to the realm of Forms have with respect to truth or falsehood as qualities of statements or beliefs about particular instances? Does the perceptible domain's lack of unmitigated truth imply that true beliefs and assertions about objects in this domain have a diminished truth value? (And what about false beliefs? Would their falsehood also be diminished?)

As far as I can tell, there are no passages in which Plato, talking about propositional truth plain and simple (and not about degrees of insight or cognitive clarity), would commit himself unmistakably to the idea that “doxastic” beliefs about perceptible objects have varying degrees of *diminished* truth value. (“Doxastic” beliefs fall outside the scope of scientific knowledge and are, hence, of an inferior kind dubbed *doxa*—non-scientific belief, “opinion”—rather than *epistēmē*.) There are, to be sure, contexts in which he characterizes such doxastic beliefs as either indeterminate with respect to truth or falsehood or as having a changing truth value. The most prominent example is the argumentation in *Rep.* V, 478e–479d. It suggests that the objects of doxastic beliefs exhibit a quasi *equivocal* mode of being oscillating between being and not being F (479b–d), which is a way of failing to be genuinely (truly) F and thus might affect the truth value of the corresponding beliefs. Yet note that this kind of argumentation has the result that none of the opposing ascriptions is more true than false; hence it supports either indeterminacy or instability,⁴³ but not the idea of definitive degrees of diminished truth value. In other passages Plato emphasizes that dialecticians, thanks to their grasp of the Forms, are able to recognize instances that reproduce a given Form in the perceptible realm, and this seems to entail that they will be able to make a judgment about such an object which is true and also justified (cf. *Rep.* VII, 520c). This sort of application of the science of Forms is, indeed, one of the main goals of Platonic philosophy since the theory of Forms is, among other things, meant to function as the foundation for ethics and normative political theory.⁴⁴ Yet I don't think that the two viewpoints exemplified by *Rep.* 478e–479d and 520c have to be incompatible. A belief may be true, and even justified, now and in a given context, yet cease to be true later or in a different context, provided that beliefs can have changing truth values, as Plato (no less than Aristotle) seems to have assumed. This allows Plato to maintain that judgments about particular

⁴³ The claim in *Rep.* 477e that *doxa* (opinion) is “not infallible”, while *epistēmē* is, should be construed as relating to the instability of truth value (cf. 479d3–5): if an object of *doxa* appears as F in one perspective, and then as not-F in another, the new appearance is a sort of falsification of the previous one. The stability of the object of a knowledge-like grasp, on the other hand, prevents such falsification (cf. 479c4, *pagiōs noēsai*). See also Vlastos (1965: 59–63); Mourelatos (2008: 129f); Szaif (2007). Compare Woodruff's comments on “strict predication” (1982: 153–6) and Burnyeat's on unqualified being and context relativity (2000: 20–2). The issue of the instability of the truth value of a doxastic belief should not be conflated with the observation that people give up their opinions relatively easily under the influence of persuasion, while knowledge remains firm (e.g. *Tim.* 51e4 and, very likely, *Men.* 98a).

⁴⁴ Other passages in the *Republic* that presuppose that we can have true beliefs even if we don't have knowledge: the discussion of civic virtues, e.g. 430b, 431d–432a; the passage on the various ways of how someone can lose a true belief, 412b–414b; the distinction of voluntary and involuntary *pseudos* (*pseudos* in the sense of “error” or in the sense of “lie”), 382aff; the explanation of how craftsmen can have correct beliefs about their products without knowledge, 601e–602a; see also 506c, 585b.

instances don't convey the *kind* of truth required for *epistēmē*, even if they result from an application of *epistēmē*.⁴⁵

But couldn't the truth value of a belief concerning a sensible object be rendered stable by adding all the necessary qualifications (time, respect, comparison, etc.)? Plato is, after all, happy to include such qualifications when he wants to defend the principle of non-contradiction.⁴⁶ And shouldn't this provide those belief contents with the kind of stable truth required for *epistēmē*? The best we can do here on Plato's behalf, I think, is to argue that the fact that a belief has to include these qualifications only shows that the truth represented in this belief is of an inferior *kind* alien to *epistēmē*. For in one way or another, these qualifications relativize the truth in question to a subjective component: the temporal position of the inquirer, the choice of a respect or comparison class, the viewing angle, etc. Platonic science looks to transcend such subjective elements of cognition.

1.2.4 Plato's analysis of propositional falsehood and truth in his later works

I have just argued that Plato understood true judgments about sensible objects as conveying an inferior, unstable kind of truth. Yet Plato's discussion of propositional truth values contained in some of his later dialogues does not hesitate to use judgments about sensible objects as examples of truth or falsehood. This needn't surprise us. For in order to set off *epistēmē*'s grasp of some firm truth against the changing and unreliable truth value of doxastic belief, Plato had to presuppose the possibility of falsehood. Yet since there were certain stock objections to the possibility of falsehood, used by practitioners of the sophistic art of disputation, he saw himself compelled to develop a theory of falsehood. Such a theory could either discuss the causes of error or analyze falsehood as a semantic property of beliefs and assertions. In Plato we find both.

Arguments that address questions concerning falsehood and truth as properties of linguistic or mental representations can be found in the *Cratylus*, the *Theaetetus*, the *Sophist*, and the *Philebus*. Among these, the discussion in the *Sophist* is the most important one for our purposes since it not only provides Plato's most detailed account of falsehood and, by implication, truth qua semantic properties, but is also the oldest extant example of such a theory. The *Philebus* contains an interesting application of Plato's

⁴⁵ Compare *Tim.* 37a–c, where the world-soul is said to have beliefs about perceptible objects (the parts of her body) that are true and reliable in all instances, yet are still *doxa* and not *epistēmē*, as opposed to the world-soul's understanding of the Forms, which is *epistēmē*. This only makes sense if we assume that the two ontological domains also represent two *kinds* of truth, one of which is inferior (which also seems to be implied in *Tim.* 51c) and not suitable for *epistēmē*.

⁴⁶ *Rep.* 436e8–437a2 does not formulate the principle of non-contradiction, strictly speaking, but rather a somewhat narrower principle according to which a thing cannot be subject to *contrary* characteristics that qualify its being, undergoing, or doing, *in the same respect, at the same time*, etc. (see also e.g. *Symp.* 211a2–5, *Parm.* 155e8–11). But this is close enough.

approach to truth and falsehood as properties of emotional states. I will have to focus here on the *Sophist* together with some brief comments on passages in the *Theaetetus* that I take to be relevant for a correct understanding of the solution in the *Sophist*.

The *Theaetetus* tackles the question of what knowledge is, but in 187d–200d the dialogue digresses to include a discussion of the possibility of error. Though highly interesting, it remains inconclusive, like most of the other arguments of the *Theaetetus*. One of the relevant points for us is how Plato describes the genesis of judgments in 189e–190a (cf. *Soph.* 263e–264a). The act of judging is (at least ideally) the endpoint of a process of deliberation in which the mind entertains a question (for instance: “Is Theaetetus beautiful?” or “What is the sum of 5 and 7?”) and weighs the pro and cons of possible answers. In the course of such a deliberation the mind tries to identify the right answer among a set of alternative possible answers. Error occurs when the mind picks a wrong alternative, believing it to be the right one. The psychological models that are subsequently introduced (the “wax-tablet” and the “aviary”) describe misjudgment as a process in which something is misidentified as a consequence of picking out the wrong mental representation from the storage of one’s memory.⁴⁷ The primary examples are instances of failed recognition of a concrete individual (as in “The person over there is Theaetetus”) and faulty mathematical equations such as “ $5 + 7 = 11$ ” (which Plato also seems to regard as attempts to *identify* a correct number as the sum of two other numbers). Yet the text also hints at the applicability of these models to straightforward predicative statements such as “Theaetetus is beautiful.”⁴⁸ But how could a predicative judgment like the one just cited be construed as involving an act of wrong identification on the *predicate* side of the judgment? In light of how Plato describes the genesis of judgment as the result of a decision between alternative answers to a question, it stands to reason that the following is his guiding idea: if Theaetetus is, in fact, beautiful, then the person who thinks that he is ugly has picked the wrong attribute from a pool of contrary aesthetic attributes and has thus misidentified Theaetetus’ actual aesthetic attribute. (The text mentions unclear perceptions and defective representations in a person’s memory as possible causes for misidentification, cf. 193c–d, 194e–195a.)⁴⁹

In the *Sophist*, the attempt to define sophists as experts of deceptive appearance is challenged by sophistic objections against the very possibility of false belief and false appearance. These objections are presented in several stages, yet I am going to mention only one argument, contained in 237c–e, because it provides the most relevant point of reference for the subsequent analysis of veridical not-being and falsehood in this dialogue. I am calling it the argument for the *impossibility of saying something which*

⁴⁷ Against Burnyeat (1990) and others, I take these two models to be an elaboration of the initial proposal to understand misjudgment as “other-judging” (*allodoxia*, 189b–c); cf. Szaif (1996: 356–93).

⁴⁸ Cf. Burnyeat (1990: 80), on the use of this example in the *Theaetetus*.

⁴⁹ We can link this approach also to an idea present in some of his earlier dialogues: Forms, once grasped, function as models whose likenesses in the empirical objects we try to recognize in our judgments (e.g. *Euphr.* 6e, *Rep.* 520c). Since successful recognition can also be described as a form of successful re-identification, errors of recognition may be considered a form of misidentification.

is not (ISNB). With an added premise, it becomes an argument for the impossibility of saying something false (ISF). The latter seems to have been a stock sophistic argument. We find versions of it in several other dialogues of Plato (*Euthd.* 283a–284a, *Crat.* 429d, *Tht.* 188c–189b). In the version of the *Sophist*, the ISNB-part of ISF is grounded on three assumptions. (For reasons of space I cannot reproduce the actual argument but only point to the role of these underlying assumptions.)

The first one, which is presented as a preliminary to all the puzzles of not-being, states that something *which is not* cannot at the same time *be*. The main speaker (an anonymous visitor from Elea) characterizes it as *Parmenides' principle* (PP):

(PP) x is not \rightarrow not: x is something which is (cf. 237a, b7–c9, 241d).

In the puzzle we are interested in, the next step builds on the assumption that whatever can be labeled a “something” also has being:

(TA) x is something $\rightarrow x$ is something which is (cf. 237c10–d11).

The text cites the contrapositive of (TA), but justifies it by pointing out, roughly, that for each x (each “something”) it is true to say that it is *one* thing, distinct from others, and that to be one thing, distinct from others, entails being.⁵⁰ The third underlying assumption (implied in 237e1–6) sets a condition for meaningful assertions:

(IA) u is an assertion \rightarrow there is (a) something which is the thing that u asserts.

(TA) and (IA) combined seem to yield the conclusion that *that which is asserted is something which is*—if we make allowances for the use of “something” as a quasi-predicate in this argument. With these two assumptions in the background, the text in 237e argues (roughly) that if (*per impossibile*) that which is asserted is *not* something which is, then the assertion asserts nothing and is, hence, not even an assertion.⁵¹ (PP) also matters for this argument since it blocks the defensive move that *that which is not* could at the same time be *something which is* and thus qualify as assertible.

From this result we get to ISF if we add another premise:

(VB) What u asserts is false \rightarrow What u asserts is something which is not.

⁵⁰ In the Greek text, the argument exploits a feature of Greek grammar, namely the fact that the word *ti* (“something”) has a grammatical *number*, the singular, as opposed to the dual or the plural. Yet I don’t think that the argument depends on this grammatical peculiarity of Greek.

⁵¹ I am simplifying the argument by leaving out intermediate steps that hinge on the etymology of the Greek word for “nothing” (*mēden*), literally: “not even one (thing)” (cf. Szaif 1996: 397f fn. 76, and on the puzzles of not-being and falsehood in general, *ibid.* 332–43, 394–412). Note that my reconstruction of the argument, which I haven’t laid out here, presupposes that “asserting x ” can function as a transparent context. For a reconstruction that treats it as an opaque context compare Wiggins (1971).

Since the phrase to the right of the arrow is vacuous (as ISNB has shown), so is the one to the left (*modus tollens*). Hence, it is not possible to assert what is false.⁵²

After this and several additional puzzles, the interlocutors conclude that they have to assail (PP) in order to vindicate the conceivability of *that which is not* and save the possibility of falsehood. An exposition of puzzles about being, introduced as a counterpart to the puzzles about not-being, serves to widen the perspective, and the dialogue then goes on to present a twofold analysis of not-being in the modes of non-identity and of negative predication (254b–9b). The upshot of this analysis is that not-being is always, in some form or other, implicitly relational, because the adverb “not” imports the notion of difference. Since, moreover, difference is a relation whose relata are things that are, only things *that are* can be something *which is not* (in the required implicitly relational sense). Thus we are no longer bound to (PP).

This strategy is then also applied to the analysis of falsehood and veridical not-being, as we will see. But why does Plato take ISNB seriously at all? And why does he not reject the move from ISNB to ISF as a simple fallacy of equivocation? It seems fairly obvious that the justification of (TA) requires a different sense of “to be” than the specifically veridical sense at work in (VB). It can’t be that Plato was simply unaware of the risk of a fallacy of equivocation, since in an earlier dialogue (*Euthydemus* 277e–278c, 295b–c) he comments on this type of fallacy.⁵³

My suggestion is, first, that he sees some genuine merit in the old Parmenidean worry about how thoughts or utterances could relate to something which does not exist. Secondly, I suspect that he also sees some initial plausibility in the claim that uttering something false amounts to uttering something which does not exist, not least because of what I have described in sections 1.1.1 and 1.1.4 as the tendency in Greek to identify an asserted truth with a piece or aspect of reality. Plato is, hence, looking for an analysis that allows him to understand how falsehood is possible without problematic reference to some inexistent entity.⁵⁴

Let’s now look at the various versions of a definition of falsehood for assertions presented in this dialogue. The first two are introduced in a passage that is still part of the puzzles concerning falsehood and not-being:

A) A statement is false *iff* it states *what is not* (= *what is not the case*).⁵⁵

⁵² This entailment is presupposed in the introduction to the puzzle in 236e1–237a10 (see Szaif 1996: 395f). It is also affirmed in 260c11–d3, as part of a transitional passage that leads up to the refutation of ISF.

⁵³ For an example of a fallacy involving the use of “is,” see *Euthd.* 283b–d.

⁵⁴ Cf. Crivelli (2012: 46–8), who defends a similar view. There is, however, much disagreement in the literature about the extent to which the concept of existence (or real existence) plays a role in the *Sophist*. But it is at least agreed that Plato rejects the Parmenidean equation of that-which-is-not with that-which-is-nothing-at-all and that he instead presents a notion of not-being that is based on the concept of difference and allows for a mixture of being and not-being.

⁵⁵ Cf. 240d9—the formulation there is actually about beliefs (*doxazein*); but in this simplified presentation I am formulating all definitional formulae for the case of assertions. Plato himself makes a transition from beliefs to assertions in 240e10.

This formulation looks odd in English but is commonsensical in Greek thanks to the specific veridical or alethic sense of “to be” (“to be the case,” “to obtain”). Yet it is immediately replaced with the following formulation (which anticipates Aristotle’s definition of propositional falsehood; see section 1.3.2):

- B) A statement is false *iff* it states *that-which-is* as *not-being*, or *that-which-is-not* as *being*.⁵⁶

The verbs “to be” and “not to be” in (B) should again be construed as instances of the veridical “to be.” (B) has greater complexity than (A) since it incorporates the distinction between affirmative and negative assertions: an affirmation states something *as being the case*, while the corresponding denial presents it as *not being the case*.

Formulation (B) still contains an unanalyzed notion of not-being *simpliciter*. The section that provides the required analysis of veridical not-being (261d–3d) starts with a discussion of two essential components of basic assertoric sentences: sequences of nouns don’t produce a sentence, nor do sequences of finite verbs, yet the combination of a noun and a finite verb does. The function of the noun in such a sentence is to name something *about which* something is said, while the verb expresses that which is said about it. The combination of these two parts of speech thus exemplifies the two basic semantic functions of picking out a subject and predicating something of it.

The dialogue then introduces two sample sentences that relate to the same subject yet are incompatible (or so they used to be before the invention of airplanes):

S₁) Theaetetus is-sitting. (*Theaitetos kathētai*.)

S₂) Theaetetus is-flying. (*Theaitetos petetai*.)

One of these two sentences (which are simple noun–verb clauses in Greek) happens to be true (since Theaetetus, who is one of the interlocutors, is sitting), while the other is false. Since (S₂) is about a real entity, Theaetetus, one can already point out as a first result that falsehood does not presuppose that the sentence be *about* something unreal or inexistent.

Yet what about the predicative part? Does it involve some not-being *simpliciter* in the case of falsehood? The concepts of sitting and flying both signify *real* natural kinds. The not-being in question can, hence, only be a function of their predicative relation to the given subject, as Plato tries to express with the following paraphrase for (S₂):

- c) S₂ states about Theaetetus that-which-is-not-with-respect-to-Theaetetus as being-with-respect-to-Theaetetus.⁵⁷

⁵⁶ Cf. 240e10–41a1. My translation is not verbatim but tries to capture the logical import of Plato’s formulation; see fn. 70 on the problems of translation. Aristotle and Plato use essentially the same grammatical construction.

⁵⁷ In this abbreviated presentation I have to streamline and simplify the sequence of formulations as we find them in the text (cf. Szaif 1996: 454–509). The formulation above is contained in 263b9. The additional qualifications have to be supplied from the context (see especially the *peri sou* in b5 and 11).

The “is” here is still veridical, yet has been relativized to the entity denoted by the subject term (comparable to the English expression “true of . . .”).

Since the two sample sentences are both affirmative, we are not given a paraphrase for the case of a negative assertion. Yet the fact that the “as being . . .”-clause is still included indicates that the distinction between affirmative and negative assertions has been retained.⁵⁸ Generalizing on Plato’s behalf, we obtain the following tentative definitional formula for falsehood in elementary assertions:

- C) A statement *s* about some object *o* is false *iff*
 (s states that-which-is-not-with-respect-to-o as being-with-respect-to-o)
 or (s states that-which-is-with-respect-to-o as not-being-with-respect-to-o).

In order to complete the analysis, the argument still has to prove that the notion of not-being involved in (C) can be derived from the concept of difference, as was the case with the types of not-being discussed in *Soph.* 254b–9b. In other words, it has to be shown that (C) is equivalent to a formulation that uses the notion of difference instead of not-being. With this goal in mind, the dialogue offers what amounts to the following description of (*S*₂):

- d₁) *S*₂ states about Theaetetus something other than that-which-is-with-respect-to-Theaetetus [as . . . ?].⁵⁹

This paraphrase does not yet supply the required “as . . .”-clause. Many interpreters, including Michael Frede (1967, 1992) and Gwil Owen (1971), read the expression “that-which-is-with-respect-to-Theaetetus” in (d₁) as referring to *everything* that obtains with respect to Theaetetus (i.e. the whole class of attributes he has).⁶⁰ If we followed this approach, we would have to expect the supplementation: “as something which is-(with-respect-to-Theaetetus).” Yet what we actually find, a few lines later in 263d1–2, is a supplementation with “as the same”.⁶¹

- d₂) *S*₂ states about Theaetetus something other (than that-which-is-with-respect-to-Theaetetus) as the same (as that-which-is with respect to Theaetetus).

⁵⁸ This was already recognized by Owen (1971: 260).

⁵⁹ This is derived from the elliptical sentence in 263b7, with some supplementations that carry over from b4–5.

⁶⁰ As I have shown in Szaif (1996: 407f, 468), based on textual parallels (see also Brown 2008: 454), the use of the neuter plural *tōn ontōn* (instead of *tou ontos*) is idiomatic in the case of a veridical usage of the participle of *einai* (“to be”) in Greek and does therefore not entail a quantification over a plurality of predicates, as the so-called Oxford interpretation inaugurated by Frede and Owen assumes (see now also Crivelli 2012: 233–59).

⁶¹ The sentence in 263d1–5 is most plausibly construed as still commenting on the sample sentence *S*₂. Most authors who have written on falsehood in the *Sophist* disregard 263d, which seems to me unacceptable since the argument of this passage ends at 263d, not 263b. The attempt by van Eck 1995 to uphold an interpretation similar to the Oxford interpretation in the face of 263d is, I think, not feasible. Crivelli (2012: 235, 245–7), has to resort to a textual emendation in order to make 263d amenable to the Oxford interpretation. See also Brown (2008: 451–61) on this whole topic.

Since Theaetetus has many characteristics, not just one, we can avoid absurdity only by assuming that the obtaining characteristic, correctly *identified* in the true assertion, is meant to fall within *a range of incompatible (i.e. contrary)*⁶² predicates such that there is exactly one predicate within the given range that is true of Theaetetus.⁶³ Accordingly, a definitional formula that captures the import of this last stage in Plato's account, replacing the "is not" in (C) with "other than," would have to include reference to incompatibility ranges.⁶⁴

Plato's decision to resort to the notion of identity at this point might simply be due to the fact that the notions of difference and identity are complementary. But I don't think that this final twist in his analysis is only such a casual, not fully thought-through move. Recall that the *Theaetetus* described predicative judgments as attempts at recognizing and re-identifying a certain feature in an object of perception or thought in response to a question that defines the relevant choice of alternatives. This points, to be sure, to a thesis about cognitive psychology and not about falsehood as a semantic property; yet Plato may have wanted to establish a correspondence between the semantic conditions of truth and falsehood and his psychology of judgment.⁶⁵

⁶² Plato has a notion of "contrary" predicates (*enantia*) that is narrower than its modern counterpart, which is why I am speaking of ranges of "incompatible" predicates (e.g. "is sitting," "is flying," "is walking," etc.).

⁶³ Versions of the incompatibility range interpretation have been defended by Ferejohn (1989) (who seems to have been the first to formulate the leading idea for this approach); Szaif (1996); Brown (2008); Gill (2009) (cf. fn. 65). This approach has to be distinguished from the proposal that the word *heteron* ("different") in 263b is used to mean "incompatible" (thus, roughly, Moravcsik 1962). The debate also involves the interpretation of the analysis of not-being and difference contained in *Soph.* 254–9. Ferejohn 1989 does not talk about the crucial passage in 263b–d in any detail, but he suggests—as does Brown (2008: 456–8)—that the analysis of negative kinds (on which cf. Lee 1972) in 257b–259b provides the basis for the implicit use of the idea of incompatibility ranges in 263B–D, while I emphasize the connection with Plato's psychology of judgment in Szaif (1996: 494ff).

⁶⁴ We would obtain the following generalized formulation (D):

A statement *s* about some object *x*, concerning some incompatibility-range *r*, is false iff [(*s* states something other than that-which-is-with-respect-to-o-within-r as being the same as that-which-is-with-respect-to-o-within-r) or (*s* states that which is the same as that-which-is-with-respect-to-o-within-r as being something other than that-which-is-with-respect-to-o-within-r)].

⁶⁵ Another (or additional) reason might be that Plato wants to establish a connection with his method of collection and division, which is very prominent in the *Sophist*. At each level of division, incompatibility ranges are discovered. To the extent that the structure of our descriptive terms reflects such taxonomies, an act of predication is always, at least implicitly, a choice among alternatives from such a range. This is how I understand the view defended in Gill (2009). Some contemporary linguistic theories also develop the idea that assertions implicitly correspond to questions, yet they do so with the aim of analyzing the semantics and pragmatics of linguistic *focus*; cf. Kadmon (2001: 261–3, 339ff).

1.3 ARISTOTLE

Propositional truth is clearly at the center of Aristotle's thinking about the concept of truth, and this is also what my comments will focus on. It should, however, be mentioned that he also talks about truth as a property of notional grasps and primary perceptual contents. In this connection, his remarks seem to point to a notion of what we might call *pre-propositional* truth or, more cautiously, *pre-judgmental* truth (since what he distinguishes it from is truth as a quality of judgment, *doxa*).⁶⁶ Note also that his *Ethics* introduces a conception of *practical truth* designed specifically for action guiding beliefs (*Nic. Eth.* VI.2, 1139a17–31).

1.3.1 The theory of the declarative sentence

Since the theory of the declarative sentence (*logos apophantikos*) in his treatise *De interpretatione* 1–6 lays the foundation for his treatment of propositional truth, I begin by recapitulating some core elements of this theory.

The basic building blocks of descriptive content and meaning are representations in our mind that have not yet been combined to form judgments. With respect to such representations, Aristotle observes that the same (kinds of) objects cause the same kinds of mental representations across the human species since the latter are “resemblances” (*homoiōmata*) that represent in virtue of a relation of similarity, presumably resulting from a natural causal process (16a3–9). The significance of phonetic signs, on the other hand, is based on conventional association with mental representations. The fact that mental representations represent in virtue of a similarity relation might suggest that the truth of beliefs is also a function of the similarity between beliefs and objects. Yet Aristotle blocks a truth-by-similarity theory by emphasizing that a mental “resemblance,” taken by itself, is not yet true or false.⁶⁷ Only with acts that set two terms or representations in relation to each other can we get to the level of judgments or utterances that are true or false (16a9–18).

⁶⁶ On the truth of primary perceptual contents (the “proper perceptibles”), cf. *De an.* II.6, 418a11–16; III.3, 427b11–14, 428a11f, 428b18–30; III.6, 430b29f; *De sensu* 4, 442b8–10; *Met.* IV.5, 1010b2f, b14–26; on the truth of notional grasps, *Met.* IX.10, 1051b17–52a4; *De an.* III.6, 430a26f, b26–30; III.7, 431a8–10. A distinct set of problems is posed by the question of how *phantasia* can be true or false (*De an.* III.3). According to Aristotle, *phantasia* can be both true or false but is not yet judgment proper, although it involves more than just an immediate sense impression. It also seems to play an essential mediating role in his cognitive psychology; cf. Caston (1996).

⁶⁷ Compare the idea discussed (but not endorsed) in Plato's *Cratylus* that a true sentence must be composed of words that are true (385b–c) and that words (morphemes) are true by way of being depictions of things (or essences) such that sentences are just more complex images (422e–424a, 424e–425a, 432e–433b).

Aristotle's next step, the analysis of the most basic constitutive parts of a sentence, owes much to Plato's *Sophist*. As in Plato's dialogue, the basic components are labeled "names" (nouns) and "verbs." We also find the same kind of conflation between the grammatical distinction of nouns and verbs as parts of speech and the logical distinction of subject- and predicate-terms.⁶⁸ A "noun" (*onoma*) has specifically the function of referring to the subject of an assertion. For if the same word serves a different function (indicated by a case ending other than the *nominative*), Aristotle no longer calls it a noun but a "noun inflection" (*ptōsis onomatos*), pointing out that it does not yield an assertion if a verb is added (16a32–b5). A "verb" (*rhēma*) has the function of signifying something *as being predicated of something else* (i.e. of the subject), which is why a verb requires a finite ending (16b6–10). Another function of the verb inflection is to indicate a time (16b8f, 16–18; cf. *De an.* III.6, 430a31f). If the predicate is expressed not by a verb, but by a noun or adjective, the sentence typically requires a connecting "... is ..." (the so-called *copula*) (16b21–5, 19b19ff; cf. *An. pr.* I.3, 25b22–4), which fulfills the same function as the verb ending in the case of verb-predicates. While sentences can have different grammatical modes, the mode of the declarative sentence (*logos apophantikos*) is singled out by the criterion that such a sentence can be true or false (16b34–17a4).

The most basic form of assertion is produced by an act in which *something is affirmed of something*, i.e. a predicate (designated by the predicate term) is asserted of a subject (designated by a noun or subject term). (Following Aristotle, I am using the expressions "subject" and "predicate" as labels for the referents of the subjects and predicate terms, viz. objects and predicated kinds, properties, etc.) The second, still elementary, kind of assertoric unit is the one that results from an act of *denying* something with respect to (literally "away from") something (17a8f, 17–22, 23–6).⁶⁹ These two basic assertoric acts are also distinguished as acts of *combining* (= affirming) and *separating* (= denying) (16a12f). "Separation," in this specific sense, is also a mode of establishing a propositional unity (*logos hen dēlōn*, 17a15f) since it is a way of setting a predicate in relation to a subject in the declarative mode.

The past and future tenses are dealt with as modifications of the predicative relation in an elementary sentence—or, what amounts to the same, as modifications of the copulative "... is ..." (16b8f, 16–18, 17a24, 29f). Quantifiers and modal operators are likewise viewed as modifications of the predicative relation. We also learn that more complex assertoric units can be formed from simple subject-predicate units when two or more such clauses are "bound together" (*syndesmos* 17a9, 16). Yet, apart from this passing and vague acknowledgment, the whole field of sentential connectives and how they affect truth value remains unexplored in Aristotle. (Note that I am using the expression "truth

⁶⁸ Cf. Barnes (2007: 93–113).

⁶⁹ On affirmative and negative predicative assertions as the two kinds of elementary assertion see also *De in.* 4, 16b26–30; *An. pr.* I.1, 24a16f, 26–8, b16–18; I.46, 52a24–38; *An. post.* I.2, 72a13f; *Cat.* 10, 12b5–16, 13a37–b3.

value” as a convenient modern term. Both Plato and Aristotle refer to truth and falsehood of sentences or beliefs as “qualities”; e.g. Plato, *Sophist*, 262e–63b, Aristotle, *Soph. el.* 22, 178b27f.)

What does Aristotle’s distinction between affirmative and negative predicative assertions amount to? We could be dealing with a distinction between affirmations and denials as two different forms of assertoric speech acts with different illocutionary force (to use the terminology of modern speech act theory). In this case, the “not” in a statement like “Socrates is not musical” would not express an element of the propositional content but rather indicate the illocutionary force of a denial. Its counterpart in the theory of belief states would be the idea that affirmative and negative beliefs are two different kinds of propositional attitude (cf. Crivelli 2004: 21f, 61f). In the context, however, of syllogistic theory (the core of Aristotle’s logic) such a distinction would be impractical. Syllogistic theory has to treat negation as part of propositional content. The situation is different if the pragmatics of speech acts or the psychology of judgment formation are taken into consideration. With this in mind, one can point out that Aristotle sees an analogy between affirming/denying and desiring/avoiding (*Eth. Nic.* VI.2, 1139a21f, *De an.* III.7, 431a8–14), which suggests that negative assertions are somehow analogous to acts of rejecting. Accordingly, we can assume that outside the context of his syllogistic, Aristotle is willing to regard denials as a distinct kind of assertoric speech act or belief attitude, which differ from affirmations in that they take the opposite stance vis-à-vis a conceivable positive state of affairs. For our purposes, however, we need not pursue this question further. I will simply speak of affirmative and negative assertions.

Affirmative and negative assertions, taken together, form pairs of “contradiction” (*antiphrasis*) (which could, again, be interpreted in a logical or in a pragmatic sense). Aristotle introduces this observation with the help of a formulation that in other texts serves him as a formula for conditions of falsehood and truth. It involves assigning the two alternate truth values to each of the two basic forms of assertion:

Since one can assert

- a) that which obtains as not obtaining,
- b) and that which does not obtain as obtaining,
- c) and that which obtains as obtaining,
- d) and that which does not obtain as not obtaining,

and likewise for the times outside the present time, it is possible both to deny whatever has been affirmed and to affirm whatever has been denied (*De in.* 6, 17a26–31).

Combinations (a) and (b) produce falsehoods, (c) and (d) truths. The Greek verb translated as “to obtain” could also mean “to belong.” It would then apply not to a complete state of affairs but more specifically to the predicate part relative to the subject part. The context of this passage suggests that this is indeed what Aristotle has in mind.

1.3.2 The definition of truth and falsehood

We encounter what has generally been regarded as the classical Aristotelian definition of truth and falsehood at the beginning of his discussion of the law of the excluded middle (LEM) in *Metaphysics* IV.7:

- G) For to assert (a) that-which-is as **not being** or (b) that-which-is-**not** as **being**, is false,
whereas to assert (c) that-**which-is** as **being** and (d) that-which-is-**not** as **not being** is true (*Met.* 1011b26f).⁷⁰

We notice again the four possible combinations that produce two instances of truth and two instances of falsehood. Thus Aristotle shares Plato's view that a definition of truth and falsehood should incorporate the distinction between affirmative and negative assertions. (The part covering falsehood is, in fact, virtually identical with the definition of falsehood in Plato's *Sophist*, 240e10f; cf. formulation (B) in section 1.2.4.) The expressions "is," and "being," printed in bold, can be interpreted as instances of the veridical "to be" (= "to be the case," "to obtain"), or as elliptical copula ("x is something or other"), or as underdetermined between the options of an elliptical copula and an existential "is." Since the construal with veridical being has the advantage of rendering the definition perfectly general, it should be the default interpretation.⁷¹ Yet it is also true that Aristotle tends to neglect complex sentences (since he has no theory for sentential connectives) and to confine his discussion to the "S is P"-scheme. Thus it is only a very small step for him from the veridical "is" to the elliptical copula.⁷²

Taken in isolation, this definitional formula is open both to a correspondence-theoretical interpretation and to minimalist readings. Tarski, in his famous article "The Semantic Conception of Truth" (1944: 342f), declares that he wants his

⁷⁰ It is not possible in English exactly to reproduce the syntactic structure of the crucial phrase in this definition (*legein to on mē einai*, etc.). Slightly more literal translations would be: "to say that that-which-is is not, etc." or "to say of that-which-is that it is not, etc." They both have their disadvantages: the first sounds as if it refers to a self-contradictory assertion, which is not what Aristotle intends. The second could suggest that the statement is about the being or not-being of the *subject*, which would force us to understand the "is" as elliptical copula or as an underdetermined elliptical-or-existential usage of "is." Yet it is perfectly possible (and even preferable for the sake of the formula's universal applicability) to read the "is" as veridical. In order to avoid the above-mentioned difficulties and to leave all options open, I am using the formulation "to assert that-which-is as not being, etc."

⁷¹ I don't think that Künne's (2003: 96f) objection against what he calls the "factual interpretation" (following Kirwan 1993) is decisive. Künne contends that this interpretation would require that a declarative sentence be *about* a fact or state of affairs. I agree with him that this would hardly be a good way of reconstructing Aristotle's theory, but I don't think that this undesirable consequence follows. An affirmation presents some state of affairs as existing/obtaining not by way of predicating *being* of a state of affairs, but *through its form* qua affirmative assertion.

⁷² This has also been emphasized by Kahn (e.g. in Kahn 1981: 112–14).

formalized semantic theory of truth “to do justice to the intuitions that adhere to the classical Aristotelian conception of truth” and goes on to quote definition G above. In section 1.3.5, I will come back to the question of what sort of truth theory Aristotle defends.

The language of *combination* (*synthesis*) and *separation* provides an alternative vocabulary used by Aristotle to formulate conditions of truth and falsehood. In *Met.* VI.4, 1027b20–23, for instance, we encounter the formulation that truth obtains if one affirms in the case of something combined or denies in the case of something separated, while falsehood results from the remaining two possibilities (see also *Met.* IX.10, 1051b3–5). This points to the following formulation of the truth conditions for affirmative and negative assertions:

- SP) “S is P” is true *iff* S and P are in a state of combination, false *iff* S and P are in a state of separation.
 “S is-not P” is true *iff* S and P are in a state of separation, false *iff* S and P are in a state of combination.⁷³

This formulation applies only to sentences that satisfy the subject-predicate scheme. Yet since Aristotle treats quantifiers and modal operators as modifications of the predicative relation, it could be adapted to quantified and modalized predications by adding adverbial specifications to the instances of combination or separation.⁷⁴

1.3.3 Temporalism and the truth-conferring relation

Aristotle firmly embraces a form of temporalism, which for my purposes here can be construed as the claim that the truth value of utterances and beliefs can change over time if they relate to things or situations that are themselves subject to change.⁷⁵

Aristotle maintains that truth and falsehood are qualities that are based on a relation, *viz.* a unidirectional “causal” (or grounding) relation that makes the truth value depend on “what is the case.” This “causal” relation between facts and truth values is,

⁷³ This formulation can be compared with Künne’s “A-Truth” in his comments on the historical development of the correspondence theory of truth (Künne 2003: 100).

⁷⁴ Thus, in order to distinguish the truth conditions of universal and particular quantifications, one would have to say that the predicate (i.e. the predicated universal) of a true universal affirmation is *combined with*, or belongs to, its subject *universally*, while the predicate of a true particular affirmation is *not universally separated from* (or does not universally not belong to) its subject; cf. *De in.* 7, 17b16–20, and compare Crivelli’s account of *De in.* 7 in 2004: 89–95. (Note that the term “separation” has to be used in order to capture the logical import of a *particular affirmation*!) While this covers existential statements of the type “Some S is P,” statements of the form “(An) S is (i.e. exists)” are often treated by Aristotle as simple predicative statements (with “is” as a predicate-verb). They can also, at least in most cases, be transformed into a predicative assertion of the form “Some . . . is S,” where the gap is filled with a description of the material substratum (thus, roughly, Menn 2008).

⁷⁵ Cf. *Cat.* 5, 4a21–b2; *Soph. el.* I.22, 178b24–9; *De an.* III.3, 428b8f; *Met.* IX.10, 1051b13–17.

obviously, not of the kind that Aristotelians call efficient causality (roughly, “causality” in the modern sense of the word). Yet a sentence is true *because* certain constellations obtain in the world, and not vice versa. Unlike efficient causality, this grounding relation can span temporal distances (at least in the case of statements about the past). The non-symmetrical character of the truth-conferring relation is emphasized in various contexts, for instance in *Met.* IX.10, 1051b6–9 and in two passages of the *Categories* (*Cat.* 5, 4b6–10; *Cat.* 12, 14b9–23).⁷⁶ The passage in *Cat.* 5 raises the question of how a sentence or belief can have contrary truth values over time. It argues that a change in truth value (or veridical quality) does not imply *real* change in the sentence or belief and, hence, also not real inherence of the veridical quality in the manner of natural properties that result from real change. A truth value is only an extrinsic, relational property, and change in truth value only a reflection of real change taking place in the external object to which the sentence or belief relates.⁷⁷ The *Cat.* 12 passage explicitly characterizes the truth-conferring relation as a “causal” or grounding relation (*aitia*) in the context of a classification of types of *priority*.

Aristotle’s views on the temporal character of truth values and the unidirectional grounding relation provide the context for his famous discussion of statements about future contingents in *De interpretatione* 9. It is clear that Aristotle’s argument there aims at preventing a deterministic consequence, the guiding intuition being that if such statements had a fixed truth value already at the time of utterance (determined by what is the case in the world), future outcomes would not be open. The question is what exactly his argumentative strategy is and what it implies regarding the principle of bivalence (PB) and law of the excluded middle (LEM). Following Weidemann and others, I am endorsing the so-called traditional interpretation (which goes back to the ancient commentaries), as it seems better to agree with Aristotle’s text.⁷⁸ It construes the argument as establishing the claim that a statement about a future contingent (e.g. a sea battle tomorrow) is not yet true or false at the present time, although it is perfectly true to assert now that tomorrow a sea battle either will take place or won’t. In other words, the distribution of the truth values among the two contradictory disjuncts is, as of now, still indeterminate in spite of the fact that the corresponding disjunction (which instantiates LEM) is valid at any given time.⁷⁹ Thus the validity of PB, understood in a temporalized fashion as the principle that *at any given time* an assertoric utterance is either true or false,

⁷⁶ See also *De in.* 9, 19a33, 18b36–19a1.

⁷⁷ Note how this contrasts with Protagoras’ theory of truth as understood by Aristotle: in *Met.* IV.6, 1011a17–20, b4–7, he points out that Protagoras’ subjectivism implies that objects don’t have intrinsic but only relational properties, since an object cannot be F *simpliciter*, but only F-for-some-believer/perceiver.

⁷⁸ On the distinction between “traditional” and “non-traditional” interpretations, see D. Frede (1985) and Weidemann (2002) (who provides the most thorough analysis of the argument that I am aware of). The most important rival interpretation claims that Aristotle only wants to refute the idea that statements about future contingents are *necessarily* true or *necessarily* false.

⁷⁹ See especially the concluding remarks in 19a27–39, and also 18b9–25. 18b17f affirms a version of LEM (p and not-p cannot both be false); and so do 19a28–32 (necessarily: p or not-p) and a35–7.

has been waived for utterances that relate to contingent events in the future, while LEM applies at all times.⁸⁰

1.3.4 The ontology of truth-bearers and truth-makers

We have seen that the truth-conferring relation holds between items that belong to the world level and items at the linguistic or mental levels. The former would be called *truth-makers* in the contemporary jargon, while the latter can be called *truth-bearers*. What kinds of entity, precisely, are truth-bearers and truth-makers in Aristotle? I'll begin with the question of truth-bearers.

First it ought to be highlighted that Aristotle nowhere commits himself explicitly to the existence of propositions in the modern sense, *viz.* as a special kind of abstract entity that serves as the primary truth-bearer and provides semantic content to both utterances and beliefs.

From *De interpretatione* 1 (16a4) we can learn that Aristotle distinguishes between spoken and written sentences but assigns priority to spoken sentences (utterances). Mental truth-bearers, in turn, enjoy some kind of priority over spoken ones insofar as language has the function of symbolizing thought in our communication with others (16a3f, cf. 23a32f, 24b1f). In his logical treatises, however, he studies the logical structure of assertoric sentences and the entailment relations among terms and sentences (e.g. *An. pr.* 24a16f, b18–20).

Mental truth-bearers could be judgments (mental events) or beliefs (dispositions), either as tokens or as types. (When we say, for instance, that two people share the same belief, we relate to a belief type.) Aristotle fails to provide us with an explicit answer as to how he wants to individuate mental truth-bearers.

As to sentences or sentence-utterances as truth-bearers, the theory of the assertoric sentence (*logos*) in *De interpretatione* 1–6 is clearly committed to treating phonetic entities (rather than inscriptions) as the primary linguistic truth-bearers (16a9–11, 19, b26, 17a17–20, 23). However, this still leaves open the choice between utterance tokens and their types. One constraint on our interpretation is that it be compatible with Aristotle's temporalism concerning truth values.⁸¹ This favors the type interpretation. The utterance type "Athens is freezing," for instance, clearly changes its truth value over time, while it would seem to have a determinate truth value if situated in a context of utterance. The token interpretation, however, tries to counter this argument by pointing out that the time index of present tense sentences (utterance tokens) can refer to an *extended present*.⁸²

⁸⁰ Cf. Weidemann (2002: 325), following von Wright (1979: 242f), on why the distinction between the validity of PB and LEM presupposes a temporalization of the truth value ascriptions. It is generally agreed that Aristotle does not intend to introduce a third truth value "indeterminate" (as promoted by Łukasiewicz).

⁸¹ See passages cited in fn. 75.

⁸² Cf. Crivelli (2004: 72–6, 183–9), who defends the utterance-token interpretation.

Other texts in Aristotle also fail to provide any really conclusive evidence to decide the issue, but generally speaking, the type interpretation seems to agree much more easily with Aristotle's temporalism concerning truth values.⁸³

After these somewhat inconclusive remarks about mental and linguistic truth-bearers, let's now address the question of what the *truth-makers* are. *Categories* 5 and 12 contain some interesting and often-quoted arguments that might point to a notion of *states of affairs*, whose obtaining or not-obtaining would then function as a truth condition. Yet on closer scrutiny, this interpretation turns out to be unsustainable or very uncertain at best.⁸⁴ The most informative texts are *Metaphysics* VI.4 and IX.10 (together with V.29). Both chapters are about the veridical (or alethic) sense of "to be" and "not to be." Recall that this sense allows for substitutivity between "that which is" and "true" (*alēthēs*). Thus, whatever turns out to be a case of veridical being, could also qualify as an instance of truth. Chapter VI.4 of the *Metaphysics* refers to states of combination (*to synkeimenon*) and separation (*to diērēmenon*) as truth conditions for affirmative and negative beliefs (1027b20–3; compare formulation SP in section 1.3.2). Yet—and this is important—it does not identify *veridical* being (or not-being) with these states of combination (or separation) at the world level, but characterizes veridical being as resulting from some condition or "affectation" of our thinking (1027b34–28a1). Hence veridical being is *not* the mind-independent truth-maker; rather it is identical with the truth value (or quality) *true* of beliefs and assertions, and this is also what we might expect in light of the mutual substitutivity of "true" and veridical "that which is (the case)."

Yet the situation becomes more complicated when we look at chapter IX.10 (together with V.29). For IX.10 does, as I am going to show, identify veridical being and not-being with the truth-makers for judgments, and this has the consequence that the truth-makers themselves also become truth-bearers and that we have to distinguish two levels of truth. First, there is what we may call the level of "objectual truth" (or world level). At this level, elementary truths are veridical states of combination between the things denoted

⁸³ Consider the example "There will be a sea battle tomorrow" in *De in.* 9. As an utterance token, it can switch the next day (i.e. retroactively) from a not yet determined truth value to the truth value *true* (or *false*). As an utterance type without a situational context, it would not have a definitive truth value at any time. Aristotle's discussion in *De in.* 9 does, I think, favor the token interpretation. But that does not mean that he has to be committed to it in other contexts. He never addresses this issue explicitly.

⁸⁴ In *Cat.* 5, 4a22–b18 and 12, 14b10–22 Aristotle twice mentions the being or not-being of a *pragma* as that which makes an utterance true or false. The word *pragma* is very vague in Greek, and it could here mean either a state of affairs or the object to which the judgment refers (i.e. its logical subject), while the "being" of the *pragma* could mean either the obtaining of a state of affairs or the being (predicational or existential) of the object that corresponds to the subject term. At first blush, the former seems more likely (cf. Nuchelmans 1973: 33–6). Yet the *Cat.* 5 passage connects the switch from being to not-being on the part of the *pragma* to notions like *kinēsis* and *pathos* (4a34–b4, b6–14) and explicitly states that only substances (roughly: self-subsistent objects) can undergo such *change* or *affectation*. This is, in fact, one of the main points of the argument in this segment. Yet states of affairs can't belong to the category of substance in Aristotle's ontology and thus cannot be subject to change in the required sense (*kinēsis*). The example used in the *Cat.* 12 (14b14–22) passage also suggests that the being in question is predicational or existential rather than veridical, hence also uses *pragma* in the sense of "object" rather than "state of affairs."

by a subject term and a predicate term (e.g. Socrates, an individual, and the property *musical*; or the *diagonal*, a universal, and the property *incommensurate*) (1051b1–5, 9–13, 17–21, 33–5). They are opposed to states of separation, which are instances of “the false” at the world level. Together, these states of combination or separation (which I will also refer to as positive or negative “constellations”) function as truth-makers for affirmations or denials respectively. Secondly, there is the level of affirmative or negative beliefs and assertions that are true or false (“semantic truth”) based on some form of agreement with truth and falsehood at the objectual level.

This distinction between objectual and semantic truth is supplemented with an analysis of (veridical) being and not-being (1051b9–13) that equates veridical being with unity (*hen einai*) (= the state of combination) between the referents of “S” and “P” and veridical not-being with plurality (*pleiō einai*) (= the state of separation). The next question, then, is how we should categorize the items to which veridical being or not-being, unity or separateness, is ascribed. I am going to speak cautiously of “subject-matters” that are constituted by pairs of items: the referents of the terms “S” and “P” (for instance an individual and a universal in the case of a singular statement). Let’s symbolize such a subject-matter as $\langle S/P \rangle$. It can be in a state of combination (Comb $\langle S/P \rangle$) or separation (Sep $\langle S/P \rangle$). (Strictly speaking, it is the components S and P that stand in the relations of combination or separation to each other.) Veridical being (= objectual truth) and not-being (= objectual falsehood) are defined as Comb $\langle S/P \rangle$ and Sep $\langle S/P \rangle$ respectively. Comb $\langle S/P \rangle$ makes the affirmative assertion “S is P” true, while Sep $\langle S/P \rangle$ renders the negative assertion “S is not P” true (semantic truth). This is equivalent to saying that the *being* of $\langle S/P \rangle$ makes “S is P” true, while the *not-being* of $\langle S/P \rangle$ renders “S is not P” true.

In light of our observations in section 1.3.1 about the modifications of the predicative relation through quantifiers and modal operators, one would expect some further elaboration on the nature and various types of these states of combination and separation. Yet chapter IX.10 only mentions the distinction between necessary and contingent states of combination or separation (1051b9–17), characterizing *necessary truths* as cases in which the subject-matter $\langle S/P \rangle$ is always and necessarily in a state of combination or unity, and *necessary falsehoods* as cases in which $\langle S/P \rangle$ is always and necessarily separated. In the case of a contingent truth, on the other hand, S and P can associate and disassociate over time, with corresponding truth value changes in the belief or assertion “S is P.” This view of modalities incorporates, hence, Aristotle’s temporalist understanding of truth values.⁸⁵

⁸⁵ What is the consequence if one of the components designated by “S” and “P” does not exist, or is not real? Restricting our scope, for simplicity’s sake, to singular and universal statements, we can be quite sure that Aristotle holds that if a fictional kind term like “goatstag” is *predicated* affirmatively (i.e. if the P-part of what is affirmed is inexistent), and the term “S” is non-empty, the resulting assertion is false. Yet how about assertions whose “S”-term is vacuous (cf. Wedin 1978)? Here we seem to get conflicting answers from Aristotle in different contexts. If we follow Aristotle’s remarks in *Cat.* 10, 13b12–35, such an affirmation counts as false, whereas the corresponding denial counts as true. Taken together with our remark about cases with a vacuous “P”-term, this would suggest that cases in which (at least?) one of the

It is tempting to interpret a subject-matter $\langle S/P \rangle$ as a “state of affairs,” and its being as the “obtaining” of a state of affairs. This is Crivelli’s approach (2004),⁸⁶ and it is also how I have occasionally (as a rough approximation) used the expression “states of affairs” in this chapter. Yet does Aristotle intend to introduce states of affairs as a distinct ontological class in his metaphysics? We should note, first, that if this is Aristotle’s intention in IX.10, there would only be *positive* states of affairs in his ontology (which is, in fact, what Crivelli claims). We would, moreover, be compelled to accept Crivelli’s conclusion that the existence of an Aristotelian state of affairs is distinct from what Aristotle in IX.10 calls its “being” (= obtaining), for the truth of a negative assertion would require that there be a positive state of affairs in a condition of *non-obtaining*. The existence of such states of affairs would be eternal (in the way abstract objects may be considered eternal), while their obtaining would be temporal (at least in the case of contingent truths). I grant that this is an attractive option from a contemporary viewpoint, yet there is no textual basis at all in IX.10 or elsewhere for such a conceptual distinction between the veridical being and the being qua existence of a state of affairs or subject-matter $\langle S/P \rangle$. We should therefore not attribute an ontology of states of affairs to Aristotle.⁸⁷

According to the interpretation I am defending, there are two crucial differences in the approach to propositional truth between *Met.* VI.4 and IX.10. First, while both texts maintain that semantic truth depends on how our affirmative or negative beliefs agree with the states of combination or separation at the world level, the latter text applies truth terminology to both levels and thus introduces a distinction between objectual

two terms of the affirmation is vacuous count as equivalent to a constellation of separation: $\text{Sep}\langle S/P \rangle$. Yet in other texts, Aristotle recognizes that there are intentional predicates that allow for true affirmative predication with respect to nonexistent subjects: for instance the predicate “opinable” (see *De in.* 11, 21a32f, cf. *Soph. el.* 5, 166b37–167a2). So here the state of combination that renders the affirmation true would *not* presuppose the existence of the subject. Another problematic case is that of purely negative (“infinite”) predicates, introduced in *De in.* 3, 16b11–15, since the existence of the subject is not required for the affirmative predication of an infinite predicate to count as true (see also *De in.* 10, 19b19–30 [cf. Weidemann 2002: 335f], *Met.* X.3, 1054b18–22).

⁸⁶ See Crivelli (2004: 3–7, 46–62) (also 2009: 90f) for Crivelli’s interpretation. (Crivelli provides the most thorough study of Aristotle’s theory of truth that I am aware of.) My own approach to *Met.* IX.10 owes much to Tugendhat (1966; 1992).

⁸⁷ It is true that the talk of composition, unity, and being *simpliciter*, which the theory of IX.10 uses in reference to a subject-matter $\langle S/P \rangle$ whenever the statement “S is P” is true, is suggestive of an ontological commitment to a new and distinct class of entities, which we might dub “veridical compounds” (corresponding to $\text{Comb}\langle S/P \rangle$) and compare to something like temporally existing positive facts. These would clearly differ from Crivelli’s eternal states of affairs, since such veridical compounds, if contingent, go in and out of existence. (They would likewise differ from propositions, because there are both true and false propositions, whereas a veridical compound would exist only as long as the corresponding affirmation is true.) It should be noted, however, that these compounds don’t have the kind of unity that constitutes an object out of parts. Rather, it is a unity of a specifically veridical character, as can be inferred from the fact that this unity obtains *only* in virtue of a predicate being true of a subject. Matthen (1983), who analyzes Aristotle’s veridical being as the existence of a specific kind of entity that he calls “predicative complexes,” runs those veridical compounds together with the being of “accidental compounds.” Yet this seems incompatible with the distinctions in *Met.* V.7 and VI.4. The theory of accidental compounds responds to a different set of problems (cf. Lewis 1991: 85ff).

and semantic truth, while the former acknowledges only semantic truth. Secondly, the latter equates veridical being with objectual truth, the former with semantic truth. This has the consequence that veridical being, as conceived in VI.4, applies also to negative assertions, *viz.* whenever they correspond to a negative constellation $\text{Sep}\langle S/P\rangle$,⁸⁸ whereas IX.10 identifies veridical *not*-being with the negative constellation $\text{Sep}\langle S/P\rangle$.⁸⁹ I hold that the view defended in IX.10 should not surprise us in the Greek context, given what we have observed about the tendency in Greek to identify a truth which is asserted, or known, with some aspect or part of reality (cf. sections 1.1.1 and 1.1.4). The innovative approach seems to be that of *Met.* VI.4, which separates veridical being from the world level.

Many modern interpreters of IX.10 wish to avoid locating truth at the world level, as an attribute of those constellations of combination that function as truth-makers for affirmations.⁹⁰ If warranted by the text, their approach would help to preserve consistency between the two chapters on veridical being in the *Metaphysics*.⁹¹ Yet in defense of my interpretation I may point out, first, that the *Metaphysics* as we have it now was put together by later editors from various drafts of Aristotle's lectures on what he called "the first [i.e. most foundational] science." The compilation is not arbitrary since it tries to follow Aristotle's own plan, but different parts were written at different times and don't necessarily always defend the same position on each issue. Moreover, notwithstanding the fact that Aristotle's formulations in the crucial passages in IX.10 (1051b1–3 and b33–5) are not as clear and unambiguous as we would like them to be, there is also *Met.* V.29, the chapter on falsehood, which draws an unmistakable distinction between, on the one hand, falsehood as quality of an account or assertion (*logos pseudēs*) and, on the other, falsehood as quality of a thing or entity at the world level (*pragma pseudos*), mentioning as a possible meaning of *pragma pseudos* the case in which something is false in virtue of being in a state of separation: for instance the pair of universals *the diagonal (of a square)* and *commensurate (to the sides)* (1024b17–21). Like V.29, IX.10 seems to link the kind of falsehood that can be equated with veridical not-being to a state of separation at the level of the *pragmata* (1051b1–5, cf. b33–5), which is logically correct only if we are

⁸⁸ See also *An. post.* I.2, 71b25f, where veridical not-being is equivalent to semantic falsehood.

⁸⁹ In *Met.* V.7, 1017a31–5 (which is the segment about "being" in the sense of "being true" in his division of the senses of "that which is"), Aristotle connects veridical being and not-being with an emphatic use of the copulative "... is ..." and "... is not ..." (cf. *An. pr.* I.37, 49a6–8, and I.46, 52a32–4). This aligns V.7 (*pace* Charles and Peramatzis 2016: 103) with V.29 and IX.10 in opposition to VI.4, for it means that V.7 correlates true negative assertions with veridical not-being, as do V.29 and IX.10. (The discrepancy is clearly recognized in Brentano 1862: 33–6.)

⁹⁰ Crivelli (2004: 52), who doesn't share it, calls it the "standard-interpretation" (e.g. Ross 1924 *ad loc.*, Burnyeat et al. 1984: 154ff); see now also Charles and Peramatzis (2016), who defend it.

⁹¹ Crivelli (2004) agrees that IX.10 acknowledges truth and falsehood at the level of *pragmata*, but he denies that this causes an inconsistency with VI.6, relying on his claim that the phrase *to de kyriōtata on alēthes ē pseudos* in 1051b1f (possibly a corrupt text) introduces a strict, technical meaning of "true" and "false," while VI.4 addresses the ordinary sense of "true" and "false" (ibid. 65f, cf. 234–7). I don't find this suggestion sufficiently plausible, not least because the talk of truth at the objectual level is grounded in the ordinary usage in ancient Greek, as I have tried to show in section 1.1.

talking about objectual falsehood, because a negative assertion expressing a state of separation can, of course, be true (semantic truth). Since it is, moreover, uncontroversial that IX.10 refers to truth and falsehood (at least) *also* as qualities of judgments, we have to conclude that IX.10, contrary to what we find in VI.4, but in line with V.29, assumes two levels of truth and falsehood such that truth and falsehood at the world level provide the truth-conditional foundation for truth as a quality of positive or negative assertions and beliefs. Recall, moreover, that the discrepancy between VI.4 and IX.10 also affects the understanding of veridical *being*, which VI.4 identifies with semantic truth, while IX.10 identifies it with the truth-makers for affirmations.⁹²

Let me conclude this section with a momentous problem that arises when we are asking what the items are that are combined or separated in an affirmative or negative belief. Aristotle very carefully distinguishes (both in *Met.* VI.4 and IX.10) between the states of combination and separation, which belong to the world level, and the linguistic or mental acts of combining and separating. It would seem that *that which is combined or separated* in these acts are not simply words or notions but rather the things (objects, properties, etc.) which they represent. For the affirmation or negation is supposed to be true if what it combines is in fact combined, or if what it separates is in fact separated (cf. *Met.* IX.10, 1051b2–5). Thus, if I, for instance, think that Socrates is (is not) musical, I am not just combining (separating) mental events, i.e. the *thoughts* of Socrates and musicality; rather I am combining (separating) Socrates and the property of musicality *in my thought*. This must not be understood in a naïve physical sense, obviously. For by combining (separating) Socrates and musicality in my thought, they do not in fact become so—otherwise the judgment could render itself true. It seems that at this point we would have to resort to a theory of intentional or representational content that distinguishes such content *both* from the mental or linguistic acts and from the corresponding items on the world-side (individuals and their kinds or properties). Yet Aristotle, unlike the Stoics after him, does not tackle this issue.⁹³

⁹² Late ancient and medieval Aristotelians have held the view that the items that exemplify veridical being are *entia rationis*, i.e. exist as intentional contents of our thinking, and are not matched one-to-one by corresponding structures of real being. VI.4 seems more congenial to this view than IX.10. There are good reasons for such a view from an Aristotelian viewpoint, for instance the problem of negated predicate terms and privations: ontologically, these terms denote the absence of something, but in the framework of a theory of truth and veridical being they are no less valid components of veridical subject-matters than their positive counterparts. The question of the status of ascriptions of passive intentional predicates to inexistent objects is also relevant in this connection.

⁹³ For Aristotle, mental representations are sensible or intellectual *forms* that are realized (“without matter”) in the mind and thus mentally reproduce forms that exist in reality, outside the mind (or are, at least, derived from other mental forms that have correlates in reality; see Charles 2000: 78ff. for a penetrating discussion of the details of this theory). On this basis, we could perhaps say that Aristotle’s remarks in VI.4 and IX.10 amount to the claim that the formation of a positive or negative belief involves the act of *representing*, in an assertoric mode, the referents of two such mental representations *as combined* or *as separated*. This is, we might say, what Aristotle means when he speaks of combining (or separating) two *external* items *in one’s thought* so as to form a belief. Yet his remarks still fall short of an *explicit theoretical* distinction of intentional content from both mental acts and real objects and their properties.

1.3.5 A correspondence theory of truth?

How does the Aristotelian definition of truth do justice to the intuition that the truth of a sentence or belief is based on some form of agreement with the facts? What makes it (if at all) a correspondence theory of truth? Traditional realist theories of truth understand truth as grounded in a relation of agreement between the truth-bearers (declarative sentences, beliefs, etc.) and the world. This idea of word-to-world agreement, as we may call it for short, has been spelled out in different ways. While modern formulations of a “correspondence theory of truth” presuppose an ontology of facts (or states of affairs), the most common way of conceptualizing word-to-world agreement in the older tradition (at least up to Kant) describes it as a relation between a thought (or its linguistic expression) and an object.⁹⁴ This view is contained, for instance, in the classical medieval formulation of truth as an “adequation” between a “thing” (or object) and a cognition (Thomas Aquinas). It goes back to similar formulations in Hellenistic and late ancient philosophy. The obvious downside of such a formulation is that it presupposes the “S is P”-scheme and cannot be applied directly to more complex sentence structures.

Aristotle’s definition of truth, like its predecessor in Plato’s *Sophist*, does not exhibit the most characteristic feature of a correspondence theory of truth, even if we grant a broader understanding of correspondence that would include the older conception of *adequation to an object*. For Aristotle does not use any term such as “correspondence” or “adequation” in his definition (see section 1.3.2). This is the reason why minimalist accounts of truth can also claim Aristotle’s definition as an ancestor.

Yet consider the formulation we find in *Met.* IX.10, 1051b4f: “. . . while someone who is in a state [i.e. a state of belief] contrary to how things are is subject to falsehood (*epseustai*).”⁹⁵ This is indicative of Aristotle’s desire to incorporate the idea of agreement or disagreement between the judgment (or the judging person) and the world in his treatment of (semantic) truth and falsehood. I think a good case can be made for the claim that the Aristotelian definition of truth, in both versions (cf. section 1.3.2), has a way of expressing the idea of word-to-world agreement *without actually using a term like “agreement.”* Let’s first take up the SP-formulation for truth and falsehood.

Any assertion that follows the “S is P”-format relates to a subject-matter <S/P> (cf. section 1.3.4) whose two components are either in a state of composition or in a state of separation. Since the assertion has to be either affirmative or negative, it is based on a choice between presenting the subject-matter <S/P> as in a state of combination or as in a state of separation. It is then obvious how there can be a match or mismatch between how <S/P> is presented and the actual state it is in. Note that it

⁹⁴ Cf. Künne’s remarks on the history of “correspondence” in 2003: 93–103, 107–11.

⁹⁵ The Greek phrase *enantiōs echein ē ta pragmata* is hard to translate verbatim. The “how” is not in the Greek text but justified by the adverbial phrase *enantiōs ē*.

is crucial here to include two alternatives to choose between. For it is the element of choice between two alternatives and the fact that such choice is, or is not, vindicated by actual combination or separation at the world level, which does the job of capturing the idea of a word-to-world agreement. This is, I submit, the reason why Aristotle does not drop the distinction between negative and affirmative assertions (or beliefs) from his definition of truth.

The same basic idea can be spelled out for formula *G*, construed as applying the notion of veridical being to subject-matters of assertions. Based on this construal, *G* is not limited to subject-matters of the <*S/P*> type but accounts for any kind of subject-matter that could be shared by an affirmation and its negative counterpart. We are, again, faced with two alternatives on the world-side (here: the veridical being versus veridical not-being of the subject-matter) and two ways in which the subject-matter can be presented, namely in the affirmative or in the negative mode. Truth-conveying agreement obtains if the being or not-being of the subject-matter matches the choice between the affirmative or negative mode of presentation.

Aristotle also embraces the idea of structural isomorphism between judgments and their truth-makers, at least with respect to sentences and judgments that are appropriately standardized (for instance by putting the term that denotes a substance in the subject position). The narrower version of his truth definition—formulation *SP*—presupposes this kind of isomorphism.⁹⁶ In this regard, Aristotle's theory of truth has much more in common with modern theories of isomorphic correspondence than with the classical post-Aristotelian account of truth as some form of *adequation* between a (non-propositional) object and the way we think or speak of it.⁹⁷

REFERENCES

- Barnes, J. (2007). *Truth, etc. Six Lectures on Ancient Logic*. Oxford: Oxford University Press.
- Boeder, H. (1959). "Der frühgriechische Wortgebrauch von Logos und Aletheia," *Archiv für Begriffsgeschichte* 4: 82–112.
- Brentano, F. (1862). *Von der mannigfachen Bedeutung des Seienden nach Aristoteles*. Freiburg i. Br.: Herder.
- Burnyeat, M. F. (1980). "Socrates and the Jury: Paradoxes in Plato's Distinction between Knowledge and True Belief," *Proceedings of the Aristotelian Society Supplementary Volumes* 54: 173–91.

⁹⁶ Cf. Crivelli's account of the mapping relation underlying "correspondence as isomorphism" (2004: 23 and 132–6). (Pace Crivelli, one should not apply the notion of isomorphism to formulation *G*, since *G* does not tell us anything about the internal structure of the truth-makers.) See also Modrak (2001: 52–62) on this topic.

⁹⁷ These comments on how Aristotle's conception of propositional truth and falsehood compares to correspondence theories of truth could also be applied to the brief sketch of a theory of falsehood in Plato's *Sophist* (see section 1.2.4).

- Burnyeat, M. F. (1990). *The Theaetetus of Plato*. Indianapolis: Hackett.
- Burnyeat, M. F. (2000). "Plato on Why Mathematics Is Good for the Soul," *Proceedings of the British Academy* 103: 1–81.
- Burnyeat, M. F. et al. (1984). *Notes on Books Eta and Theta of Aristotle's Metaphysics*. Public Manuscript, Oxford: Sub-Faculty of Philosophy.
- Brown, L. (1994). "The Verb 'to be' in Ancient Greek Philosophy: Some Remarks," in S. Everson (ed.), *Language [= Companions to Ancient Thought 3]*. Cambridge: Cambridge University Press, 212–36.
- Brown, L. (2008). "The *Sophist* on Statements, Predication, and Falsehood," in G. Fine (ed.), *Oxford Handbook of Plato*. Oxford: Oxford University Press, 437–62.
- Caston, V. (1996). "Why Aristotle Needs Imagination," *Phronesis* 41: 20–55.
- Charles, D. (2000). *Aristotle on Meaning and Essence*. Oxford: Oxford University Press.
- Charles, D. and M. Peramatzis (2016). "Aristotle on Truth-Bearers," *Oxford Studies in Ancient Philosophy* 50: 101–41.
- Code, A. (1993). "Vlastos on a Metaphysical Paradox," in T. Irwin and M. C. Nussbaum (eds.), *Virtue, Love, and Form: Essays in Memory of Gregory Vlastos [=Apeiron 26]*. Edmonton: Academic Print. & Publ., 85–98.
- Crivelli, P. (2004). *Aristotle on Truth*. Cambridge: Cambridge University Press.
- Crivelli, P. (2009). "Aristotle on Signification and Truth," in G. Anagnostopoulos (ed.), *A Companion to Aristotle*. Chichester: Wiley-Blackwell, 81–100.
- Crivelli, P. (2012). *Plato's Account of Falsehood: A Study of the Sophist*. Cambridge: Cambridge University Press.
- Eck, J. van (1995). "Falsity without Negative Predication," *Phronesis* 40: 20–47.
- Enders, M. and J. Szaif (eds.) (2006). *Der philosophische Wahrheitsbegriff in seiner Geschichte*. Berlin: de Gruyter.
- Ferejohn, M. T. (1989). "Plato and Aristotle on Negative Predication and Semantic Fragmentation," *Archiv für Geschichte der Philosophie* 71: 257–82.
- Fine, G. (1979). "False Belief in the *Theaetetus*," *Phronesis* 24: 70–80.
- Frede, D. (1985). "The Sea-Battle Reconsidered," *Oxford Studies in Ancient Philosophy* 3: 31–87.
- Frede, M. (1967). *Prädikation und Existenzaussage*. Göttingen: Vandenhoeck.
- Frede, M. (1992). "The *Sophist* on False Statements," in R. Kraut (ed.), *The Cambridge Companion to Plato*. Cambridge: Cambridge University Press, 397–424.
- Gill, M. L. (2009). "Method and Metaphysics in Plato's *Sophist* and *Statesman*," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2009 Edition), <http://plato.stanford.edu/entries/platosophstate/>.
- Heitsch, E. (1962). "Die nicht-philosophische *alêtheia*," *Hermes* 90: 24–33.
- Hestir, B. (2004). "Plato and the Split Personality of Ontological *Alêtheia*," *Apeiron* 37: 109–50.
- Hestir, B. (2016). *Plato on the Metaphysical Foundation of Meaning and Truth*. Cambridge: Cambridge University Press.
- Irwin, T. H. (1977). "Plato's Heracleiteanism," *Philosophical Quarterly* 27: 1–13.
- Kadmon, N. (2001). *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*. Malden, MA: Blackwell.
- Kahn, Ch. H. (1973/2003). *The Verb "Be" in Ancient Greek*. Dordrecht: Reidel [reprinted by Hackett, 2003, with a new introduction].
- Kahn, Ch. H. (1978). "Linguistic Relativism and the Greek Project of Ontology," in M. Sprung (ed.), *The Question of Being*. University Park: Pennsylvania State University Press, 31–44.
- Kahn, Ch. H. (1981). "Some Philosophical Uses of 'to be' in Plato," *Phronesis* 26: 105–34.

- Kirwan, C. (1993). *Aristotle. Metaphysics, Books Γ, Δ, and E*. Oxford: Oxford University Press.
- Krischer, T. (1965). "ETYMOΣ und ΑΛΗΘΗΣ," *Philologus* 109: 161–74.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Lee, E. (1972). "Plato on Negation and Not-Being in the *Sophist*," *The Philosophical Review* 81: 267–304.
- Lee, M.-K. (2005). *Epistemology after Protagoras*. Oxford: Oxford University Press.
- Lewis, F. (1991). *Substance and Predication in Aristotle*. Cambridge: Cambridge University Press.
- Luther, W. (1935). "Wahrheit" und "Lüge" im ältesten Griechentum. Borna-Leipzig: Noske.
- Matthen, M. (1983). "Greek Ontology and the 'Is' of Truth," *Phronesis* 28: 113–35.
- Menn, S. (2008). *Aristotle on the Many Senses of Being* (manuscript).
- Modrak, D. (2001). *Aristotle's Theory of Language and Meaning*. Cambridge: Cambridge University Press.
- Moravcsik, J. M. E. (1962). "Being and Meaning in the *Sophist*," *Acta Philosophica Fennica* 14: 23–78.
- Mourelatos, A. (2008). *The Route of Parmenides*, revised edn. Las Vegas: Parmenides-Publishing.
- Nehamas, A. (1975). "Plato on the Imperfection of the Sensible World," in *Virtues of Authenticity*. Princeton: Princeton University Press (1999) (reprinted from *American Philosophical Quarterly* 12: 105–17).
- Nehamas, A. (1982). "Participation and Predication in Plato's Later Thought," in *Virtues of Authenticity*. Princeton: Princeton University Press (1999) (reprinted from *Review of Metaphysics* 36: 343–74).
- Nuchelmans, G. (1973). *Theories of Proposition*. Amsterdam: North-Holland Publishing.
- Owen, G. E. L. (1971). "Plato on Not-Being," in G. Vlastos (ed.), *Plato*, vol. 1. Garden City (NY): Anchor Books, 223–67.
- Ross, W. D. (1924). *Aristotle's Metaphysics (Text, Introduction, Commentary)*, 2 vols. Oxford: Clarendon Press.
- Santas, G. (2001). *Goodness and Justice: Plato, Aristotle, and the Moderns*. Malden, MA: Blackwell.
- Szaif, J. (1996). *Platons Begriff der Wahrheit*. Freiburg/Munich: Alber (2nd edn, 1998).
- Szaif, J. (2004). "Die Aletheia in Platons Tugendlehre," in M. van Ackeren (ed.), *Platon Verstehen. Perspektiven der Forschung*. Darmstadt: WBG, 183–209.
- Szaif, J. (2007). "Doxa and Epistêmê as Modes of Acquaintance in *Republic V*," in *Les Études platoniciennes*, vol. 4. Paris: Les Belles Lettres, 253–72.
- Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341–75.
- Taylor, C. C. W. (1999). *The Atomists Leucippus and Democritus*. Toronto: University of Toronto Press.
- Tugendhat, E. (1966). Review of "K. Oehler, *Die Lehre vom noetischen und dianoetischen Denken bei Platon und Aristoteles*," *Gnomon* 48: 752–60.
- Tugendhat, E. (1992). *Der Wahrheitsbegriff bei Aristoteles*, in *Philosophische Aufsätze*. Frankfurt/M.: Suhrkamp, 251–60.
- Vlastos, G. (1965). *Degrees of Reality in Plato*, in *Platonic Studies*, 2nd edn. Princeton: Princeton University Press (1981): 58–75 (reprinted from R. Bambrough (ed.), *New Essays on Plato and Aristotle*. New York: Humanities Press, 1–19).

- Vlastos, G. (1966). *A Metaphysical Paradox*, in *Platonic Studies*, 2nd edn. Princeton: Princeton University Press (1981): 43–57 (reprinted from *Proceedings and Addresses of the American Philosophical Association* 39: 5–19).
- Wedin, M. (1978). “Aristotle on the Existential Import of Singular Sentences,” *Phronesis* 23: 179–96.
- Weidemann, H. (2002). *Aristoteles: Peri Hermeneias* [= *Aristoteles: Werke*, ed. H. Flashar, vol. 1, pt. II], 2nd edn. Berlin: Akademie Vlg.
- Wiggins, D. (1971). “Sentence Meaning, Negation, and Plato’s Problem of Non-Being,” in G. Vlastos (ed.), *Plato*, vol. 1. Garden City, NY: Anchor Books, 268–303.
- Woodruff, P. (1982). *Plato: Hippias Major. Translated, with Commentary and Essays*. Oxford: Blackwell.
- Wright, G. H. von (1979). “Time, Truth, and Necessity,” in C. Diamond and J. Teichman (eds.), *Intention and Intentionality*. Ithaca, NY: Cornell University Press, 237–50.

CHAPTER 2

TRUTH IN THE MIDDLE AGES

MARGARET CAMERON

2.1 INTRODUCTION

PHILOSOPHERS throughout the medieval tradition displayed tremendous interest in truth: in its character, whether humans can know it (in this lifetime and/or beyond), how humans can know it, whether there are different kinds of it, and so on. To get a handle on the various debates over, and theories about, truth during this period, we must first appreciate the extent to which medieval theories were rooted in ancient resources, primarily in the ideas of Aristotle and Augustine. In section 2.2, we will review these ancient conceptions of truth and their uptake in the medieval period, primarily in the thirteenth and early fourteenth centuries. Given the various conceptions of truth, there was intense debate over how truth is cognized, and in section 2.3 we will examine some of the major developments and arguments. Many debates over the nature of truth and how we cognize it took place in what is to us an unfamiliar venue, namely, in the study of Aristotelian logic, which served as the foundation for education. It is not surprising, then, that a new theory of truth emerged out of a distinctive logical tradition known as “terminism,” or the “logic of the moderns.” This new semantic theory of truth will be surveyed in section 2.4.

2.2 CONCEPTIONS OF TRUTH: THE ANCIENT RESOURCES

It is impossible to understand medieval theories of truth without appreciating the ancient resources on which they relied. Medieval authors held up various ancient figures as authorities, in both their theological and philosophical writings. Two authors are most important here: Aristotle and Augustine. From Augustine (and also from

Scripture), medieval thinkers inherited a rich metaphysical understanding of truth. From Aristotle, they inherited two theories that would be critical for the development of theorizing about truth: first, that truth claims are propositional, and second, that the highest truth, *scientia*, is the conclusion of a deductive syllogism. In this section we will look at each of these conceptions of truth in turn, briefly illustrating them with one or two examples of their medieval uptake.

2.2.1 The metaphysical status of truth

The metaphysical, or ontological, status of truth may strike contemporary readers as a category mistake in a volume on truth, and it is certainly the conception of truth that has the least currency today. But is a characterization stemming from the work of Plato, and accordingly has a deep and enduring pedigree. In *Republic* VI, Plato provides the very well-known description of the relationship between the Good and the other Forms, in which the Good shines its light onto the Forms for their intelligibility.¹ Augustine picks up this imagery of light and develops it into a master narrative in which God is shining upon our minds the light of eternal truth. Augustine's imagery is further supported by Christian Scripture, including the well-known passage in the Gospel of John according to which Christ (God) is truth and light.² According to this metaphysical conception of truth, humans are wholly dependent upon God for understanding eternal and unchanging truth. Thus, from this conception of truth as God there developed the theory of cognition known as the doctrine of illumination (see section 2.3.1).

According to Augustine,

You cannot deny the existence of an unchangeable truth that contains everything that is unchangeably true. And you cannot claim that this truth is yours or mine or anyone else's; it is present and reveals itself in common to all who discern what is unchangeably true, like a light that is public and yet strangely hidden.³

The truth is identical to God, Augustine argues, since God alone is eternal and necessary, and this is what it is to be true. This view seems to imply that there is only one thing that is true, thereby ruling out, for example, contingent truths. It also raises the question: given that there can be, on this view, true things, are all true things identical with God? These matters were avidly discussed in the medieval tradition.

The metaphysical status of truth—the view that *things* are true—had several consequences for medieval philosophy. Many medieval thinkers held the view that truth is one of the transcendentals, which means that truth transcends all of the categorical

¹ Plato, *Republic* VI, 508e–510a.

² John 14:6.

³ Augustine, *On Free Choice of the Will* II.12, trans. Williams.

ways of being (e.g. being as substance, quantity, quality, relation, as well as Aristotle's other categories).⁴ As a transcendental, anything that has being, then, is true; this idea is expressed by claiming that truth is *convertible with* being. In the article on "Truth" in the *Summa theologiae*, Thomas Aquinas (1225–74) claims that, just as artifacts are called true to the extent that they correspond to the design in an artisan's mind, so natural things are true to the extent that they correspond to the species in God's mind: "for instance, a rock is called a true rock when it attains the proper nature of a rock as preconceived in God's mind."⁵ Citing Augustine as his key authority, Aquinas also appeals to the Arabic philosopher Avicenna (c. 980–1037) ("The truth of a thing is a property of its *esse* [being] that is stable for it"), as well as Anselm of Canterbury (1033–1109). Given the extent of Anselm's authority over efforts to vindicate ontological truth, a few words on his views are needed.

Anselm, best known for his ontological proof of God's existence, wrote expressly in his treatise *Monologion* that God is truth.⁶ In his dialogue *On Truth* (*De veritate*), Anselm converts this claim into a question, explicitly reflecting one of the concerns we mentioned regarding Augustine's ontological notion of truth as God:

Since we believe that God is truth *and* we say that truth is in many other things, I wish to know whether whenever truth is spoken of we ought to be saying that it is God of whom we speak.⁷

To answer, he first sets out to determine whether there is truth in speech, by which he means in affirming or denying that something exists. Anselm's answer is negative, since he holds that a statement is only true when it *participates in* the truth—for that is when the statement signifies *what it ought*, Anselm explains. In this way, Anselm assimilates truth to rectitude (rightness), and claims that

if truth and rectitude are in the essence of things because they are that which they are in the highest truth [i.e. in God], it is certain that the truth of things is rectitude.⁸

Truth is, according to Anselm (and quoted by Aquinas), "a rectitude perceptible by the mind alone."⁹ Thus, in answer to the question, "whether whenever truth is spoken of we ought to be saying that it is God of whom we speak," it is so; and our judgment that something is true is a claim that what is in reality is aligned with, or rectitudinal with, God's mind.

⁴ The sense of "transcendental" here is nothing like Kant's. Rather, it is synonymous with "transcategorical."

⁵ Aquinas, *ST* Part I, question 16. This comes from Aquinas's own opinion regarding the question, "Does truth exist in things or only in the intellect?"

⁶ Anselm, *Monologion*, ch. 18.

⁷ Anselm, *De veritate*, ch. 1, trans. McInerny; emphasis added.

⁸ Anselm, *De veritate*, ch. 7, trans. McInerny.

⁹ Anselm, *De veritate*, ch. 11, trans. McInerny.

But being perceptible by the mind alone should not be taken to mean that truth is a mental phenomenon, on this view. Many thirteenth-century masters held the view that truth should not be dependent upon knowers. Rather, they appealed to Augustine's claim that, "The true is that which is," which Aquinas characterized as, "The true is the indivision of being and that which is."¹⁰ However, this ontological conception of truth was not compatible with Aristotle's views, even if medieval masters could find support for it here and there among Aristotle's writings. But from the philosopher who held that "falsity and truth are not in things—it is not as if the good were true, and the bad were in itself false—but *in thought*,"¹¹ there arose a markedly different conception of truth.

2.2.2 The propositional status of truth

Aristotle's logic formed the basis of education in the medieval curriculum. One of his treatises, which was continuously available to Latin scholars and which began to be studied in earnest at the turn of the twelfth century in the West, was *On Interpretation*. This was taken to be the second of Aristotle's logical treatises, to be read just after reading *Categories*. *On Interpretation* provides its readers with the basic components required to make propositions: the name/noun and the verb. These two components can be divided (e.g. "Cats do not walk," dividing walking from cats) or combined (e.g. "Cats walk"), to yield true or false claims. According to Aristotle,

However, just as some thoughts are in the soul without being true or false, yet some are necessarily one or the other, so too with spoken expressions. For composition and division have to do with falsity and truth. Hence, names and verbs themselves are similar to thoughts without composition and division, such as "human" or "white", when nothing further is added; for neither is true or false at this point. An indication of this: for "goatstag" signifies something, but is not yet true or false, if neither "to be" [i.e. is] or "not to be" [i.e. is not] is added, either simply or in accordance with time.¹²

Aristotle excludes truth and falsity from other types of verbal expression, such as words taken individually and prayers,¹³ to which Boethius (c. 475–c. 526) (in his commentary on *On Interpretation*) also adds a list of other types of phrase, such as commands and wishes.¹⁴ According to Boethius, "in reality, [affirmative and negative propositions] are

¹⁰ Aquinas, *De veritate* I, 1, trans. Mulligan. For Aquinas's views on this definition of truth, see Aertsen 1992.

¹¹ Aristotle, *Metaphysics* IV, 4 1027b25–6, trans. Ross; emphasis added.

¹² Aristotle, *On Int.* 16a10–18, according to Boethius's Latin translation (*Peri hermeneias* [vel *De interpretatione*], A.L. II.1–2).

¹³ Aristotle, *On Int.* 17a1–6.

¹⁴ Boethius was a fifth-century commentator on Aristotle's philosophy. He was responsible for translating and transmitting the first of the logical treatises into the Latin tradition.

equal, for they participate equally in truth and falsity.”¹⁵ On this basis, medieval thinkers inherited a propositional theory of truth.¹⁶

As the bearers of truth and falsity, propositions are used in syllogistic reasoning, the topic taken by most medieval scholars to be at the core of the program of study of logic. Since it would be anachronistic to talk in terms of a medieval *epistemology*, it ought rather to be said that propositional truth and falsity were treated in *logic*, which during the thirteenth century can be described as cognitive theory, that is, as investigating and describing the cognitive processes that enable humans to access truth and detect falsity.¹⁷ Cognition of reality is mediated by the senses and the intellect, such that on their views direct access to reality itself was not possible. As we will see in section 2.3, there were extensive debates during the medieval period over how this process works.

Throughout the thirteenth century, truth was defined in terms of an adequation between reality and the intellect:

The first appearance of being to the intellect is: being agrees (*concordet*) with the intellect. Indeed, the agreement is called the “adequation of the intellect and reality (*rei*)”, and in this conformity is perfected the principle of the true. This is therefore what the true adds to being, namely, conformity or adequation of reality and the intellect. As it is said, regarding this conformity, cognition follows reality. Therefore, the being of reality (*entitas rei*) precedes the principle of truth, but cognition is an effect of truth.¹⁸

While this is the best-known definition of truth from our period, it is in fact difficult to say precisely what adequation consists in. Thomas Aquinas glosses it as “conformity,” and we ought to take this gloss literally and appreciate that this theory of truth is conformalist: the form of a thing in reality *just is* the form of the thing in the human intellect.

From among the countless examples of medieval discussion and debate over this definition of truth, let us look at the rich theory of propositional truth given by the early fourteenth-century philosopher Walter Burley (c. 1275–c. 1344). A brief note about the theory of signification, or signs, at work here needs to be made first. Medieval scholars drew from Aristotle’s so-called “semantic triad” (properly, a tetrad) to explain the order and relation between those elements involved in linguistic signification:

Thus those which are in vocal expressions are signs (*notae*) of passions of the soul, and those which are written are signs of those which are in vocal expression. And just

¹⁵ Boethius, *Peri herm.* II, 18.

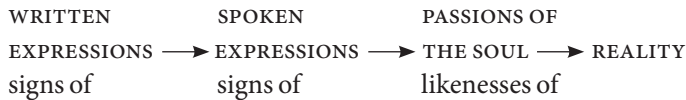
¹⁶ The language of “participating in truth” is Neoplatonic, and it should be noted that medieval thinkers, although firmly rooted in the Aristotelian corpus, read it through the lenses of many authoritative sources who were influenced by, if not part of, the ancient school of Plato. On the theories of the medieval proposition (*propositio*), see Kretzmann 1970 and Nuchelmans 1973.

¹⁷ There was another tradition of logic in the thirteenth century (indeed, with roots in the twelfth) that served as the basis for the terminist logical tradition; see section 2.4.

¹⁸ Thomas Aquinas, *De veritate*, q. 1, a. 1 (5.162–6.200). This definition of truth appears on a list with several other definitions of truth in *Summa de bono* by Philip the Chancellor (c. 1160–1236) as the definition of truth “of the philosophers.” See Pouillon (1939).

as written expressions are not the same for all, so neither are vocal expressions the same; however, of those which are of the first signs, the passions of the soul are the same for everyone, and those which are likenesses (*similitudines*) of reality itself are also the same.¹⁹

This passage itself has generated much controversy, but for now we will use it just to appreciate the order it sets out:



This semiotic chain is grounded in reality itself, such that the ultimate significate of individual words is ontological. There was a lot of debate over what Aristotle meant by “passions of the soul,” and Walter Burley mentions at least four options circulating at the time:

- (i) the intelligible species, or what we may call a “concept”;
- (ii) an act of knowing;
- (iii) a term, or end-point (*terminus*) of an act of knowing, that is, something formed by the intellect as an image or likeness of the extramental thing; or
- (iv) the extramental thing *itself* insofar as it has the ability to move the intellect.

At the level of propositions, in which truth and falsity are found, the same type of chain-ing structure is identified between different types of propositions. Burley recognized five: written, spoken, those that are subjectively in the intellect, those that are objectively in the intellect, and—interestingly—real (*propositio in re*).²⁰ Their order is this:

The written proposition signifies the spoken proposition, the spoken proposition signifies the conceptual proposition, namely the proposition composed out of concepts. I ask, now, whether the proposition composed out of concepts is the ultimate significate, namely what does not signify anything further, or whether it signifies something further. The former cannot be granted, because the concepts, out of which the proposition in the mind is composed, signify. Therefore, the entire proposition composed out of concepts signifies.²¹

Thus, written propositions are signs of spoken ones, and spoken ones are signs of mental ones. The mental proposition, too, signifies: it signifies a *real proposition*. Burley

¹⁹ Aristotle, *On Int.* 16a3–9, from Boethius’s translation.

²⁰ There is ongoing controversy over how to understand this doctrine, and whether Burley held two different theories of the proposition over his career. See Cesalli (2007); Conti (2011).

²¹ Walter Burley, *In Cat.* Prologue, fol. c3vb–c4ra (quoted in Cesalli 2013: 124n).

scholars have debated just what this could be, yet it seems best to consider it as a *fact* (so, as something known by the mind, rather than, say, a state of affairs in the world). This is because the real proposition is not real in the sense of being wholly extramental. Rather, by distinguishing between propositions that are in the intellect either objectively or subjectively, Burley has devised a way to anchor the chain of propositions in reality insofar as it exists *objectively* in the mind: rather than the proposition being merely an accidental feature of the intellect (as an accident in a substance), the proposition is objective in the sense that it takes as its *object* extramental reality itself. In other words, what Burley takes to be the matter of the *propositio in re* (as opposed to its form, which is the copula indicating the relation between its terms) is “totally outside the intellect,”²² insofar as the reality signified by the subject and predicate terms are *in* reality. In this sense truth is the adequation of the intellect *to* reality.²³ By terminating the chain of signifying propositions in a proposition which has objective being, Burley avoids the problem of infinite regress. It also avoids the problem raised by Burley himself, namely, if a true mental proposition has only subjective being, that is, it is an accidental feature of a person’s intellect, then everyone would have their own truths, so to speak!²⁴

To be sure, Burley’s explanation of propositional truth adds a great deal of ontological baggage. It makes propositions into real things, after all. As we will see in section 2.4, this and other types of cognitively rich theories of truth will be subjected to Ockham’s razor, resulting in a new, lighter type of semantic truth.

2.2.3 The paradigm of scientific knowledge

A third major influence on medieval theories of truth comes from Aristotle’s *Posterior Analytics*, which began to be used in the early to mid-thirteenth century as part of the logical curriculum. In this treatise, Aristotle famously sets out the paradigm for what he calls *episteme*, which is translated (somewhat misleadingly, given the fact that the word is overdetermined by historical forces) into Latin as *scientia*. Rather than transliterating this term as “science,” it is better to construe it as scientific *knowledge*, and recognize it as advocating what has been called an “epistemic ideal.”²⁵

According to *Posterior Analytics*, *episteme* (or *scientia*) is described thus:

We think we understand a thing *simpliciter* (and not in the sophistic fashion accidentally) whenever we think we are aware both that the explanation because of which the object is, is its explanation, and that it is not possible for this to be otherwise. It is

²² Walter Burley, *In Cat.* Prologue, fol. c4rb (quoted in Cesalli 2013: 129n).

²³ Compare this with Aquinas’s formulation of the adequation of intellect *and* reality, which does not indicate a direction-of-fit.

²⁴ Walter Burley, *Commentarius in lib. Periherm.* section 1.25, 61. I am especially grateful to the work on this topic by Cesalli (2007; 2013).

²⁵ Recently, Pasnau (2014) has been advocating for a return to this ideal to be used in contemporary epistemology.

clear, then, that to understand is something of this sort . . . Hence that of which there is understanding *simpliciter* cannot be otherwise.²⁶

From this passage medieval thinkers recognized the first two of four requirements for something to count as *scientia*:

- (i) it requires knowledge of the cause(s) of what is known; and
- (ii) it must be the type of truth that is necessary, that is “not possible . . . to be otherwise.”

Aristotle added:

[W]e know through [syllogistic] demonstration. By demonstration I mean a scientific deduction; and by scientific I mean one in virtue of which, by having it, we understand something.²⁷

This added the third requirement for *scientia*:

- (iii) its conclusion is entailed by deduction, that is by a demonstrative syllogism.

A fourth requirement—not found in Aristotle in any way—was appended by medieval thinkers:

- (iv) it must be known with *certainty* and *evidentness*.

For those unfamiliar with the medieval philosophical tradition, it might seem a surprise that the concern with certainty was at play, since this epistemic criterion is so heavily associated with René Descartes. But the rationalist orientation that was pervasive in medieval theology and philosophy raised the requirement for *scientia* to include the demand for certainty.

It may appear extreme that the only type of truth countenanced by this paradigm is necessary truth. However, if we read some of the medieval writing on this topic, it becomes apparent that knowledge of necessary truths was an epistemic goal, with the assistance of which other types of truths could be identified. The identification of these other types of truth, which turn out, in fact, to be gradations of truth (e.g. contingent, testimonial, etc., as we will see), demonstrate clearly that Aristotle’s *scientia* was held by medieval thinkers to be an *ideal*.

Consider for example William of Ockham (c. 1287–1347), who in his commentary on Aristotle’s *Physics* notes that “there are different meanings of ‘*scientia*,’ and one is not

²⁶ Aristotle, *Post. An.* 71a7–16, trans. Ross.

²⁷ Aristotle, *Post. An.* 71a18–19, trans. Ross.

subordinated to the other.”²⁸ Given the apparently strict requirements for *scientia*, it is a bit of a shock to see Ockham’s first meaning of the term:

In one sense, *scientia* is certain cognition of something that is true. In this sense, some truths are known only on trust; for instance, when we say we know that Rome is a big city, although we have not seen Rome. In the same way I say I know that this person is my father and this person is my mother. And so with many other things that are not “evidently” known. And yet because we adhere to these things without a shadow of doubt, and they are true, we are said to know them.²⁹

These are truths one knows on the basis of trust—presumably the trust in another person’s epistemic character and their testimony. The second sense of *scientia* is knowledge of *contingent* facts: “For instance, if no one told me that the wall is white, I should still know that the wall is white, just by seeing whiteness on the wall.”³⁰ The third and fourth types are of increasing strictness in keeping with Aristotle’s own criteria:

In a third sense, *scientia* means an evident cognition of some necessary truth. In this sense, no contingent facts are known, but only first principles and the conclusions that follow from them. In a fourth sense, *scientia* means an evident cognition of some necessary truth caused by the evident cognition of necessary premises and a process of syllogistic reasoning.³¹

Ockham supplies his readers with a few other senses, but the point is clear enough now. Aristotle’s claim that knowledge *as an ideal* properly consists in necessary truths, whose causes are known and which entail a demonstrative conclusion, is one type of truth, to which the medieval thinkers added an even tighter requirement, namely that it be known with certainty and evidentness. But it would be impractical and absurd to restrict truth to just this one type, and most medieval thinkers recognized gradations of knowledge along with gradations of types of truth (i.e. necessary, contingent, testimonial). Since everything in the so-called created realm is subject to change and destruction, it would seem that this idealized truth was nearly impossible to achieve.³²

A note needs to be made here regarding the requirements that true propositions be known with certainty and evidentness. A very clear guide on these criteria is

²⁸ Ockham, *Expositio super viii libros Physicorum*, 4, trans. Boehner. Ockham is a nominalist in the senses that (a) he believes that only individual, not universal, things are real, and (b) he investigates—even in a commentary on Aristotle’s *Physics*—not extramental reality but only the meanings of our words.

²⁹ Ockham, *Expositio super viii libros Physicorum*, 4–5, trans. Boehner.

³⁰ Ockham, *Expositio super viii libros Physicorum*, 5, trans. Boehner.

³¹ Ockham, *Expositio super viii libros Physicorum*, 5, trans. Boehner.

³² It is interesting to note that mathematical knowledge was excluded from being *scientia* in the strict sense, since *scientia* requires knowledge of its subjects’ causes, but the axioms on which mathematics is based are not causes of mathematics.

John Buridan (c. 1300–60) in his *Summulae de Dialectica*, in one of its treatises that concerns demonstrative knowledge. Buridan states unequivocally that knowledge requires that true propositions be known with certainty and evidentness. Certainty requires two things: that the proposition assented to be true, and that one's assent be firm, that is, without doubt or fear.³³ Buridan explains what is meant by "evidentness" by contrasting it with "that credulity that we believers ought to have concerning the articles of the Catholic faith e.g. that God is triune."³⁴ The belief that God is triune, Buridan admits, has certainty (because he takes this belief to be true), but it lacks evidentness.

Intriguingly, Buridan then entertains two skeptical hypotheses that make clear the vulnerability we as cognizers experience as possible knowers of truth.³⁵ Buridan recounts this scenario raised against the view that knowledge about natural and moral truth is possible:³⁶

For example, we could not know that the sky is moving, that the sun is bright and that fire is hot, because these are not evident. God could annihilate all these, and it is not evident to you whether they exist. Or God could put the sky to rest or remove light from the sun or heat from fire. And finally, they say that it is not evident to you concerning the stone you see as white that it is such that it is white, for even without the whiteness and the stone God can create in your eye an image [*species*] entirely similar to the one you have now from the object; and thus you would make the same judgment as you do now, namely, that there is white stone here.³⁷

This judgment would be false, Buridan points out, ruling out both certainty and evidentness. In response, Buridan (like Ockham) acknowledges that perfect certainty and evidentness is not always possible, and that they obviously admit of degrees. As a corollary to this sceptical scenario, Buridan countenances not only the scenario described above in which God—without our knowing—changes the way the world is, but also one in which he plays with our knowledge itself: "supernaturally it is possible for my [act of] knowledge, while it remains the same, to be converted into non-knowledge." Regarding this, Buridan confesses that it is not a requirement that our purportedly true beliefs "cannot be falsified by any power" whatsoever.

³³ Buridan, *Summ. de dialectica* section 8.4.4, trans. Klima.

³⁴ Buridan, *Summ. de dialectica* section 8.4.4, trans. Klima.

³⁵ On medieval treatments of epistemic skepticism, see Lagerlund (2010). We find little evidence of the sort of Pyrrhonic, global skeptical worries that would grip Descartes and others in the early modern era. It has, of course, been long recognized that Descartes's ideas did not emerge from him entirely *de novo*. The epistemic turn that is said to occur with Descartes was, it seems, one in which the problem of skepticism came to *drive* the philosophical agenda and its concern with the acquisition of truth and certainty.

³⁶ Klima (2001: 708n) credits the controversial philosopher Nicholas of Autrecourt (c. 1295–1369) with this skeptical scenario.

³⁷ Buridan, *Summ. de dialectica* section 8.4.4, trans. Klima.

2.3 COGNIZING TRUTH: DEBATES IN COGNITIVE PSYCHOLOGY

The literature on medieval debates over, and theories about, cognizing truth is massive, complex, and still very much being investigated today, in many cases for the first time since it was originally written. Medieval thinkers recognized the ancient idea that the human intellect is the highest and best part of being human; they also recognized that the perfection of the human intellect was a means to be closer to God. However, arriving at a theory of cognition that could meet the incredibly high standards set by Aristotelian *scientia*, accommodate the Aristotelian theory that all cognition begins in sense experience, and not violate either the authority of their ancient resources or religious belief, proved to be an arduous task. In this section we will outline some of the main lines of debate over how to cognize truth that took place in the thirteenth and early fourteenth centuries, just before the introduction of a new (semantic) theory of truth. Since nearly every philosophically-minded theologian and theologically-minded philosopher weighed in on the topic, we have here chosen representatives and highlights of this exciting period of philosophical debate. First, we will look at the widely held belief that cognition occurred by divine illumination, a view taken from Augustine, using Bonaventure of Bagnoregio (c. 1217–74), better known as St Bonaventure, as our representative of this way of arriving at truth. Second, the impact of the views of Thomas Aquinas on subsequent (and even contemporary) philosophy is enormous, not excluding this topic. He is a representative of the adequation theory of truth, a view typically (but mistakenly) held to be “the” medieval theory of truth. Aquinas saw no need for illumination and argued a naturalist line. Third, and in response to Aquinas’s effort to eliminate the need for illumination, Henry of Ghent (c. 1217–93) aimed to synthesize Aristotle and Augustine by re-introducing the need for illumination in cognition. Fourth, and finally, we will look at critical arguments given by John Duns Scotus (1265/6–1308) against Henry, and see how Scotus ushers in both a new way of thinking about cognition and our grasp of truth, and a crisis for theories of truth based on these types of cognitive theory.

2.3.1 Cognition by illumination—Bonaventure

The idea that the human mind is illumined by God has its origins in the Platonic tradition, and for figures in the medieval tradition the idea gained traction largely because of the influence of Augustine. In outline, the various versions of the doctrine of illumination held that, on its own, human cognition is incapable of grasping the truth, since the truth is eternal and immutable, whereas humans are mortal and corruptible. The language of “illumination” replaces Plato’s language of “recollection,” that is, Plato’s view that all learning is recollection.³⁸ It is important to appreciate, however, that by the

³⁸ A recollection doctrine would be heretical to those with religious faith who believed that every soul is created anew by God, and thus, that there is no metempsychosis.

time Aristotle's entire corpus was once again made available to masters in medieval universities in Europe in the thirteenth century, his works pertaining to truth and cognition were being read by thinkers who were already committed to the role of divine illumination. Intriguingly, efforts to blend aspects of Aristotle's theories with an illuminationist perspective were prevalent (such as in the very first medieval commentary written on Aristotle's *Posterior Analytics* by Robert Grosseteste (c. 1168–1253)).³⁹

Bonaventure's commitment to illuminationism can be found throughout his writings, and he is an excellent example of a figure sufficiently familiar with Aristotelian philosophy who is yet prepared to promote illuminationism using some of Aristotle's own terminology. Consider his answer to this question found in his *Disputed Questions on the Knowledge of Christ*: "Whether that which is known by us with certitude is known in the eternal reasons themselves."⁴⁰ Ultimately Bonaventure will argue affirmatively, but not before providing thirty-four arguments *pro* and twenty-six arguments *contra*. The majority of arguments *pro* are simply quotations taken from the authority of Augustine and other Church Fathers. But Bonaventure, being familiar with Aristotle's philosophy—frequently borrowing from it and adapting it as needed—also cites Aristotle as an authority for illuminationism!⁴¹ Bonaventure's positive philosophical (as opposed to authoritative) arguments are structured in a similar way: contrasting the certainty, infallibility, unlimitedness of immutable truth to the mutable, fallible, limited human mind, he argues that the light of eternal reasons is what enables knowledge to be acquired with certainty. The negative arguments come in large part from Aristotle, whose theory of knowledge is rooted in sense experience, proof of which (according to Aristotle) is that loss of one of the senses results in loss of knowledge of things having to do with that sense. Moreover,

[t]he principles of being and knowing are identical [according to Aristotle]. Therefore, if the principles of being that are proper and intrinsic to the creatures themselves are nothing other than created principles, then whatever is known is known through created reasons and not, therefore, through eternal reasons and lights.⁴²

The list of arguments *contra* is long, but it is clear that the Augustinian-inspired Bonaventure feels much pressure from Aristotelian cognitive theory.

Bonaventure's positive answer to the question is this:

[The eternal light] is not the sole principle of knowledge, nor is it attained in its clarity; but together with the proper *created* reason it is known obscurely and as in a mirror.⁴³

³⁹ See Van Dyke (2009).

⁴⁰ Bonaventure, *On the Knowledge of Christ* Q 4, trans. Hayes.

⁴¹ Bonaventure cites *Nicomachean Ethics*, Book 3 for Aristotle's description of knowledge: "Therefore, the object of scientific knowledge is necessarily eternal, for all absolutely necessary beings are eternal. And eternal things are ungenerated and incorruptible." Bonaventure concludes, "Therefore, there can be no such thing as certain knowledge unless the very nature of eternal truth is involved. But this is found only in the eternal reasons. Therefore, etc."

⁴² Bonaventure, *On the Knowledge of Christ* Q 4, trans. Hayes.

⁴³ Bonaventure, *On the Knowledge of Christ* Q 4, trans. Hayes.

There are three ways to interpret this answer, as Bonaventure explains. The first, Platonic, interpretation takes God's eternal light as the "total and sole cause of knowledge," but this cannot be correct because then all knowledge would be revelation, which is clearly false. The second is that it is only the influence of, but not the eternal light itself, that causes reason. Bonaventure rejects this out of hand simply on the authority of Augustine, who disagrees with the idea that by its own action human minds have some access to the eternal light. The third, and acceptable, interpretation makes God into the moving (efficient) cause of human knowledge of universals and first principles. As E. Gilson characterized Bonaventure's position, "The demands of truth are absolute, since its distinctive characteristics are those of God; no one will not attain to it who does not attain to God."⁴⁴ But as a consequence of original sin and humanity's subsequent "fallen nature," this knowledge is not in perfect conformity with the divine light, and this is why Bonaventure characterizes knowledge of truth as "obscure and dark."

2.3.2 Cognition by intelligible species—Thomas Aquinas

Much has been written on Aquinas's theory of cognition. Here we are more narrowly interested in his challenge to the illuminationists: that illumination is not required for knowledge, and a naturalist account is available. We will also introduce the role played in theories of cognition by intelligible species—that is, what are better known today as "concepts."

We have already seen in section 2.2.2 the theory of truth as the adequation of intellect and reality, which is at the heart of Aquinas's explanation of how humans cognize the truth. Aquinas held that "[e]very cognition occurs through some *species* of the cognized thing in the cognizer."⁴⁵ A *species* is variously glossed as a form, a likeness (*similitudo*), or a representation (*repraesentatio*). Following Aristotle's doctrine, Aquinas recognized that a cognition is a change undergone by the cognizer by means of a reception of a form (species) in the intellect.⁴⁶ As a result, the cognizer and what is cognized are formally identical; that is, the cognizer is assimilated to, or is likened to, the object cognized, and the form of that object is present in the cognizer. Of course, receiving the form of something, say, turtles, in the intellect does not make the intellect into a turtle—rather, the form of turtle is in the intellect as known. This is what Aquinas thinks of as intellectual cognition, and it is facilitated, as we will see, by means of the intelligible species. Aquinas also recognizes another type of cognition, namely sensory cognition, since when forms are received by any of the five senses (touch, taste, sight, hearing, smell), those sense organs are changed by that reception. In this case, it is a sensible, not an intellectual, species that gives rise to cognition of particular things (e.g. this cat), whereas

⁴⁴ Gilson (1965: 356).

⁴⁵ Thomas Aquinas, *In Sent.* 1.36.2.2 *sed contra*.

⁴⁶ Thomas Aquinas, *ST* 1.75.5.

intelligible species enable cognition of universals and generalities (e.g. cats, felinity).⁴⁷ According to Aquinas, it is not the intelligible species itself that are known. Rather, it is by way of (*id quo cognoscimus*) intelligible species that universals—that is, universal truths—can be known.⁴⁸ If a truth becomes known, it is what Aquinas variously called a concept (*conceptus*), mental word (*verbum interius*), or intention as understood (*intentio intellectus*).⁴⁹

To appreciate Aquinas's argument for the rejection of the role of divine illumination, we will need to get a sense of the intelligible species, since this is what does the work that God's light was thought to do.⁵⁰ The intelligible species is that *by which*, Aquinas contends, things can be known. He argues against the view that intelligible species is that which is known with the following reasons:

First, the things we have intellective understanding of are the same things that scientific knowledge is about. Therefore, if the things we have intellective understanding of were just the species that exist in the soul, then it would follow that all types of scientific knowledge are not about things that exist outside the soul, but only about the intelligible species that exist within the soul—just as, according to the Platonists, all types of scientific knowledge are about the ideas, which they claimed to be the things which are actually understood.⁵¹

Second, Aquinas argues, it would follow that every proposition, including contradictory ones, would be simultaneously true. This is because intelligible species cannot exist extramentally (precisely *because* they are intelligible, not material), and,

if a power has cognition only of what it receives within itself, then that is all it makes judgments about. But something seems to be the case insofar as the cognitive power is affected in a given way. Therefore, the cognitive power's judgment will always be a judgment about what it itself receives as such; and so every judgment will be true.⁵²

Consequently, intelligible species must be that by which something is known. This can be thought of as a kind of intellectual tool, already present in the intellect, by which the information, which has been gathered in the cognizer's imagination on the basis of the sensory species it has received, can be intellectually grasped. This intellectual tool, then,

⁴⁷ Thomas Aquinas, *ST* 1.12.4.

⁴⁸ Thomas Aquinas, *ST* 1.85.2.

⁴⁹ Intentions are equivalent to concepts. Medieval authors recognized two types: first intentions, which are cognitions of particular things ("cat," "tree"), and second intentions, which are second-order cognitions of types of first intentions (e.g. "species," "genus"). On the recent debates over how to interpret Aquinas on concepts and mental words, see Panaccio (1992); Pasnau (1997); Perler (2000); Stump (2003); Brower and Brower-Toland (2008).

⁵⁰ There is also considerable debate over the interpretation of Aquinas's doctrine of intelligible species, for which see fn. 50 and references contained in those sources.

⁵¹ Thomas Aquinas, *ST* 1.85.2, trans. Freddoso.

⁵² Thomas Aquinas, *ST* 1.85.2, trans. Freddoso.

is part of the (immaterial) structure of the intellect. (As we will see in section 2.4, this part of Aquinas's doctrine came under attack by William of Ockham, which resulted in the doctrine's eventual demise.) As a tool the intelligible species need to be set into act; that is, existing potentially in the intellect, they need to be actualized in some manner. How is this possible? Rather than positing that it is the work of something supernatural, Aquinas postulates that it is the work of the agent intellect. Drawing on Aristotle's doctrine in *De anima*, Aquinas recognized that human beings have both a passive and an active intellect. The precise nature of the agent intellect was a subject of tremendous controversy, indeed leading to charges of heresy and worse.⁵³ These details do not need to detain us here, however, since the important point is that Aquinas has found a way to naturalize the process of cognition. Thus, there is no need for illumination from God to cognize universal truth, since we are naturally capable of doing so given our human, rational nature.

Aquinas's theory was very influential on later thirteenth-century philosophy. However, the elimination of the role of the divine in our access to truth did not sit well with everyone, as we will see in the case of Henry of Ghent.

2.3.3 Synthesizing Aristotle and Augustine—Henry of Ghent

Henry of Ghent was well aware of the clashing views of the two great authorities in the thirteenth century on the question of human cognition of truth. For Aristotle, all intellectual knowledge is based in sense experience of particulars, whereas for Augustine, "the pure truth is not to be sought from the senses of the body."⁵⁴ Henry's grand move was to attempt to synthesize their positions by accommodating some form of natural human knowledge alongside the doctrine of divine illumination. To do so, he introduced a key distinction between *the true* and *the truth*. He acknowledged that, by taking knowledge in its widest sense to include sensitive cognition—the cognition of individuals by means of the senses—"it must be said without qualification and absolutely that it is possible to know and apprehend something with certain sense knowledge."⁵⁵ However, the case is different and more complex for intellectual knowledge—knowledge of first principles and of universals—and relies on the true/truth distinction.

⁵³ Aquinas himself was caught up in this controversy. Against the Arabic philosopher Averroes (Ibn Rushd; 1126–98), who held that the agent intellect was a separate intellect and not a faculty of the human soul, Aquinas wrote *On the Unicity of the Intellect against the Averroists*.

⁵⁴ Augustine, *On Eighty-three Questions*, Q 9. Henry of Ghent's views are subtle and reflect ongoing revision throughout his lengthy career. On the differences, see Marrone (1985). In this section we will focus on just one period of his writing, taken from his *Summa of Ordinary Questions*, which he began to write in 1276.

⁵⁵ Henry of Ghent, *Summa* art. 1, Q 2, trans. Teske.

Henry relies on the commonplace doctrine of the convertibility of truth and being, such that anything that has being is true, and the true has being. To know by means of the intellect that something exists, or has being, is to know it is true. This can be done by means of an act of simple apprehension:

For every cognitive power that by its knowledge apprehends a thing, as it has being in itself outside the knower, apprehends what is true in it, but it does not by this apprehend its truth. For even in animals, a sense correctly apprehends about a thing what is true in it, such as a true human being, a true tree, a true stone, and especially its proper object, concerning which it is necessarily true, but it still apprehends or knows the truth of nothing.⁵⁶

The idea here is that just as sense experience in non-rational animals permits their awareness of true beings outside of themselves, so simple intellective cognition in rational animals permits knowledge of true beings outside of themselves. But to grasp the truth about something requires a more complex order of intellectual operation, namely, to compose and divide what is first grasped by simple intellection of incomplex particulars (e.g. “horse,” “runs”). In this sense, Henry’s position is in alignment with the Aristotelian doctrine that there is truth only in composite or divided sentences (e.g. “A horse runs/does not run”).

Henry describes the activity of composing and dividing as making judgments: “to apprehend by a certain judgement about a thing that in the truth of the thing it is *such* or *such*.” The question, which Henry explicitly puts to himself, is whether this can be done naturally or whether it requires divine illumination. Following the definition of truth as the conformity of the intellect and the thing, Henry rightly recognizes that this definition requires a relation between the mental and the extramental object, and in order to determine whether that relation is conformal there needs to be comparison with what he calls an “exemplar.” Exemplars come in different types, and Henry distinguishes two exemplars in the created world, and a third in the divine (i.e. God’s own ideas). One type of created exemplar is, say, a picture of some person. Supposing one had never met the person depicted, it would be possible on the basis of the exemplar only to have a mere imaginative apprehension of the thing. While it is likely that, upon meeting that person, one would be able to identify him on the basis of the imaginative apprehension, this does not meet the criteria for knowledge of the truth, since it is fallible and uncertain. The second type of created exemplar is, putting the terminology in more familiar terms, the concept in the human mind acquired on the basis of abstraction from sense experiences. Henry is quick to point out that both Aristotle and Augustine recognize this sort of exemplar and its role in knowledge-making. However, it too fails to meet the criteria for knowledge of the truth of things, because the concept has been generated on the basis of a changeable world, and is thus vulnerable to this impermanence; moreover, the human mind is itself fallible. Raising a third reason against this created

⁵⁶ Henry of Ghent, *Summa* art. 1, Q 2, trans. Teske.

exemplar as a source for truth, Henry introduces a skeptical doubt: “For through the same images of sensible things in a dream or madness, we judge that the images are the things themselves, and when awake and sane, we judge that [same thing] concerning the things themselves.”⁵⁷ As a consequence, knowledge of truth that meets the criteria for *scientia* requires that it be seen “only in relation to the eternal exemplar.” In the end, Henry has found room for both some natural knowledge (i.e. sense knowledge, which is attenuated and imperfect) and divine illumination.

It should be noted that Henry does not supply his readers with many details about the mechanics of divine illumination. This is frustrating, and the mystery of the relation between the divine light and the human mind remains. But Henry does explain how the divine light enables judgments of truth regarding the conformity of intellect and thing: it illumines the quiddity, or essence, of extramental things to allow for comparison with what is in the mind. In other words, a thing’s *truth*, or what Henry called its “pure truth” (*sincera veritas*) is its *essence*.

2.3.4 Intelligible species without illumination—John Duns Scotus

Against Henry’s renewed efforts to preserve a role for divine illumination, John Duns Scotus argues away the need for it, and attempts to defend the naturalness of human knowledge by explaining the need for intelligible species to explain cognition. Recall that medieval authors distinguished between sensible and intelligible species, or concepts, as defended by Aquinas in his account of human cognition. The question facing Scotus was whether intelligible species were both necessary and sufficient for explaining intellectual cognition.⁵⁸

Considering narrowly Scotus’s arguments against Henry on the topic of illumination and the role of the intelligible species, we need first to explain an important feature of Scotus’s philosophy: the difference between what he called the formal and the real distinction.⁵⁹ In addition, for any given thing in reality, it must be recognized that it has what Scotus calls two “aspects”: on the one hand, it has a singular aspect, since it exists as a particular thing; on the other hand, it has a universal aspect, since it is a *kind* of thing that has a common nature. Thus, any given tree, say, this individual pine tree, has a singular aspect since it is distinguished from other trees by its distinguishing difference (called a *differentia*). But it also has a common or universal aspect insofar as it has a nature that is common with other trees of its type. According to Scotus, the distinguishing difference and the common nature are *formally distinct*, by which he means that

⁵⁷ Henry of Ghent, *Summa* art. 1, Q 2, trans. Teske.

⁵⁸ For an overview on debates over intelligible species, see Tachau (1988); Spruit (1993).

⁵⁹ For an overview of Scotus on the formal distinction, see McCord Adams (1982).

although we can in theory separate them, one never finds in nature either a common nature or a distinguishing difference on its own.

This individual entity is not matter or form of the composite, inasmuch as each of these is a [common] nature. Rather, it is the ultimate reality of the being that is matter or that is form or that is the composite. Thus whatever is common and yet determinable can still be distinguished (no matter how much it is one thing) into several formally distinct realities of which this one is not formally that one. This one is formally the entity of singularity and that one is formally the entity of the [common] nature. These two realities cannot be distinguished as *thing* and *thing* . . . Rather when in the same thing, whether in a part or in the whole, they are always formally distinct realities of the same thing.⁶⁰

In contrast, any particular quality of a given tree is *really distinct* from other qualities. For example, the pine tree's having *x* number of needles, since the greenness of the tree is a *real distinction* from its number of needles.

With these distinctions in mind (i.e. the two aspects of things and the formal and real distinction), Scotus takes aim at Henry's theory insofar as Henry attempted to replace the role of the intelligible species (and thereby the agent intellect) with the light of divine illumination. On Henry's theory, somehow the divine light must be sufficient to render the sense experiences collected in the human imagination into a known, essential fact, and Scotus complains that it is difficult to see how this is possible. Instead, Scotus argues, the agent intellect must act on the phantasm—the product of the imagination which has stripped away by a process of abstraction the distinguishing features of some thing: for example, abstracting away the particular number of needles of this pine tree and that one, and so on—in order to produce an intellectual cognition of the thing itself. To do so requires the role of the intelligible species, which is able to give rise in the intellect to knowledge of the thing according to its universal aspect. Neither the imagination on its own, nor the illumination by God, is capable of performing this act; they are the wrong sorts of “tools,” so to speak, to perform the cognitive work. Moreover, as Scotus points out,

If conformity to an uncreated exemplar is meant [by Henry], conformity to it cannot be recognized unless the exemplar itself is known, for unless the term of a relation is known the relation itself cannot be known. Consequently, it is false to assume that an eternal exemplar is the reason why we know something when this exemplar itself remains unknown.⁶¹

As with Aquinas, Scotus recognized that the intelligible species, not God's light, is that *by means of which* intellectual knowledge is produced.

⁶⁰ Duns Scotus, *Ord.* II, d. 3, p. 1, qq. 5–6, n. 188, trans. Spade 1994, 107, with slight modification.

⁶¹ Duns Scotus, *Ord.* I, dist. 3, q. iv, art. IV, trans. Wolter.

Scotus's argumentative strategy is to show that Henry's theory of cognition is mistaken, that his arguments "lead to skepticism," and also that, somehow, Henry misunderstood Augustine's own theory. The latter strategy is aimed not merely at preserving Augustine's enormous authority (which Scotus deeply accepted), but at trying to demonstrate the compatibility of Augustine's views with his own theory. Scotus's success on this front is debatable. However, it seems that after Scotus there was no more room for illuminationism, although soon enough there would be too no more room for intelligible species in the explanation of the human cognition of truth.

2.4 SEMANTIC TRUTH: THE NEW PARADIGM

William of Ockham characteristically applies his razor, that is, his principle of parsimony, to the thirteenth-century theories of cognition: common natures and intelligible species are eliminated altogether. Ockham builds upon a logical tradition known as "terminism," since it takes as the subject-matter of logic linguistic and mental terms, in which major logical developments and new theoretical tools were applied. On this basis Ockham promotes a radically new theory of truth: semantic truth. Although it is very much in keeping with Aristotle's own logical theory of truth, the so-called "modern" logical developments allow for a greater precision about matters of truth, as well as for a radical de-ontologization of the topic as a whole.

2.4.1 Ockham's critique of previous theories of cognition

First, let us briefly look at Ockham's arguments against the dominant thirteenth-century theories of cognition. Recall that, for Aquinas and his followers, humans cognize intellectual truth by means of intelligible species, which can be described as the intellectual tool that enables the generation of a concept or mental word. The intelligible species play the role that was played by the divine light in theories of illumination. Ockham, however, dispenses with intelligible species altogether. Ockham found it intolerable that intelligible *species* are, on the Thomistic view, prior to intellectual cognitive acts.⁶² On Aquinas's view, there are the following items involved in intellectual cognition: the intelligible species, the agent intellect, which moves, or actualizes, the intelligible species, and the concept or mental word that is thereby generated. These three phases are superfluous, according to Ockham, since he holds that cognition can be accounted for by positing only intellectual acts and what he calls "habits." In other words, thinking is just a matter of engaging in individual intellectual acts. When a person is either sleeping or not actively thinking (at all, or about something in particular), the intellect can still

⁶² William of Ockham, *Rep.* II, q. 12–13, *OTh* V, 253; see Panaccio (2004: 28–9).

be said to retain the content of those acts by means of intellectual habits that have been generated on the basis of prior thinking.⁶³

One question stands out: how, if Ockham has dispensed with intelligible species, which were the very tool used by the intellect in order to attain universal truth, can humans attain universal truth with only extramental objects and mental acts? In one swift move, Ockham answers this question: there is no universality in the world, no common natures or the like which need to be cognized by intellectual means. Instead, universality is only a feature of the intellect—a matter of how we think about things, how things are signified by us—and nothing more.⁶⁴ This is what Ockham's nominalism amounts to: namely, that there are only individuals in reality. Let us now turn to his semantic, or logical, theory of truth to see how he supports this new philosophical orientation.

2.4.2 Truth as co-supposition

As we have seen, Ockham pared down the complex thirteenth-century intellectual architecture, leaving only intellectual acts and habits. He referred to these intellectual acts as “mental signs,” or “mental words.” The mental words can be combined or divided into mental sentences which are then truth-evaluable. However, they are not truth-evaluable by determining whether they are adequate to reality, as was the case for earlier philosophers. Instead, they are truth-evaluable by ascertaining whether the terms in the subject and predicate positions in those mental (and spoken)⁶⁵ sentences *supposit* for the same thing. To appreciate Ockham's theory of truth, we will need to review the doctrine of supposition.

The doctrine of supposition is an indigenously medieval logical theory that has no precedent in any ancient resources, including Aristotle's. As we have seen, Aristotle did provide his readers with a schema, the semantic triad, that accounts for the relations between spoken and written words, cognitions or understanding, and reality. To signify, according to most medieval authors, was to give rise to an understanding of something in someone's mind. A word's signification was recognized to be one of its properties (i.e. it is a psycho-causal property).⁶⁶ But medieval logicians began to identify other

⁶³ There is ongoing debate over the extent to which Ockham's theory is more akin to Aquinas's than either he or some of his proponents would prefer: see Pasnau (1997); Panaccio (2004). Also relevant to this theory are the two types of intellection developed by Ockham: intuitive and abstractive.

⁶⁴ William of Ockham, *SL* I, 14, trans. Spade 1974.

⁶⁵ William of Ockham, *SL* I, 3, trans. Spade 1974, is primarily interested in mental signs, and he points out significant differences between mental and spoken words: they “differ in that although all of the grammatical features of mental names belong to spoken names, the reverse is not true . . . Case and number belong to mental and spoken names alike . . . On the other hand, gender and declension are grammatical features peculiar to spoken and written names.” Ockham's language of thought is universal. The best study of this topic is Panaccio (1999).

⁶⁶ This description is from Spade (1982).

properties of terms (*proprietates terminorum*), including the property of supposition.⁶⁷ According to Ockham, “[s]upposition is a property of a term, but only when it is in a proposition.”⁶⁸ That is,

Supposition is said to be a sort of taking the place of another. Thus, when a term stands for something in a proposition in such a way that we use the term for the thing and the term or its nominative case (if it is in an oblique case) is truly predicated of the thing (or the pronoun referring to the thing), the term supposits for that thing; or this, at least, is true when the term is taken significatively.⁶⁹

Terms have signification, and except when they are being taken to refer to themselves (e.g. “‘Human’ is a noun”), in which case they have *material supposition*, they do not lose this signification. They give rise to an understanding of something in someone’s intellect. However, when words are combined into propositions, they often behave in ways that cannot be explained by their signification alone. Types of supposition, which are other types of property that words used in propositions can acquire, were distinguished to account for role in the proposition as a whole. Take the straightforward case of material supposition mentioned earlier. In the sentence, “‘Human’ is a noun,” the subject-term “Human” supposits for, or *stands for*, a name, and not for what is signified by the name. Ockham recognized, in addition to material supposition, two other types: personal and simple supposition.

Personal supposition occurs, Ockham explains,

when a term supposits for the thing it signifies, whether that thing be an entity outside the soul, a spoken word, an intention [i.e. concept] of the soul, a written word, or any other thing imaginable. Thus, whenever the subject or predicate of a proposition supposits for its significatum so that it is taken significatively, we always have personal supposition.⁷⁰

More simply put, a term that has personal supposition *stands for* whatever it signifies. Keeping in mind that, for Ockham, the extramental world consists of only particular, individual things, a term can never personally supposit for an extramental common nature, or universal *in re*, since these things do not exist in his sparse ontology. How, then, can we construe the supposition of “animal” in the proposition, “Dogs are [a species of] animal” if no species of animal exists in extramental reality? For Ockham, “animal” supposits for the individual animals themselves.

⁶⁷ On the origins of supposition, see De Rijk (1963–67); Bos (2013), and the entire volume of the journal *Vivarium* 51 (2013).

⁶⁸ William of Ockham, *SL I*, 63, trans. Spade 1984.

⁶⁹ William of Ockham, *SL I*, 63, trans. Spade 1984.

⁷⁰ William of Ockham, *SL I*, 64, trans. Spade 1984.

The other type of supposition recognized by Ockham is simple supposition:

Simple supposition occurs when a term supposits for an intention [i.e. a concept] of the soul and is *not* functioning significantly. For example, in “*Man is a species*,” the term “man” supposits for an intention of the soul, for *it is that intention* which is a species. Nonetheless, the term “man” does not, properly speaking, signify that intention.⁷¹

What Ockham means by simple supposition is just that because, in this case, the term “species” has no extramental referent, since there are no species in reality, the term “man” stands for the concept “species.” We can now begin to see how this theory is used by Ockham as a theory of truth.

In the treatise on propositions in his *Summa logicae*, Ockham states that the truth of a proposition consists in the co-supposition of its subject and predicate terms. In its first introduction, Ockham gives what is perhaps a confusing example (confusing, at least, to us):

[F]or the truth of “This is an angel” it is not required that the common term “angel” be *really identical with* what is posited as the subject, or that it *be really in* that subject, or anything of that sort. Rather, it is sufficient and necessary that the subject and predicate *supposit for the same thing*.⁷²

In this case, the demonstrative pronoun “this” and the general term “angel” supposit for, or stand for, one and the same thing. If so, then the proposition is true. Since Ockham has ruled out the possibility that there are common natures or universals in reality, we are not left looking for some general referent of the term “angel.” Let us use one other example to make the co-supposition theory of truth clearer, this time using a proposition that turns out to be false. Ockham considers the following universal proposition: “Every animal is healthy.”

[A]ll propositions like the following are, properly speaking, false: “Every animal is healthy,” assuming that one lion is healthy and one ox and one man, and so on; “Every animal was on Noah’s ark”; and so on for many others. For they have many false singulars, and it is not the case that the predicate agrees with all those things for which the subject supposits.⁷³

Thus, given that the subject “every animal” supposits for this animal and that animal and that animal, and so on, there are inevitably going to be cases in which one of the supposits turns out not to be healthy, and so the proposition, “Every animal is healthy”—given the way individual animals are in reality—will never be true.

⁷¹ William of Ockham, *SL I*, 64, trans. Spade 1984.

⁷² William of Ockham, *SL II*, 2, trans. Freddoso 1998; emphasis in original.

⁷³ William of Ockham, *SL II*, 4, trans. Freddoso 1998.

What is exciting about Ockham's theory of truth as co-supposition is that it cuts out altogether any need to account for human cognition when accounting for a proposition's truth. It is in this sense that it is a semantic, or perhaps we could say, logical, theory of truth.⁷⁴ All that is required is to identify, given the propositional context of use of the subject and predicate terms, whether they both stand for, or supposit for, one and the same thing.

REFERENCES

- Aertsen, J. (1992). "Truth as Transcendental in Thomas Aquinas," *Topoi* 11: 159–71.
- Anselm of Canterbury (1996). *Monologion and Proslogion, with the Replies of Gaunilo and Anselm*, trans. T. Williams. Indianapolis: Hackett.
- Anselm of Canterbury (1998). *On Truth/De veritate*, trans. R. McInerny, in B. Davies and G. Evans (eds.), *Anselm of Canterbury: The Major Works*. Oxford: Oxford University Press, 151–74.
- Aristotle (1965). *On Interpretation*, in Boethius (trans.), *Peri hermeneias [vel De interpretatione]*, *Aristoteles Latinus* II.1–2, ed. L. Minio-Paluello and G. Verbeke. Turnhout: Brepols, 5–38.
- Aristotle (1984). *Metaphysics*, trans. W. D. Ross, in J. Barnes (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation*, vol. 2. Princeton: Princeton University Press.
- Aristotle (1984). *Posterior Analytics*, trans. W. D. Ross, in J. Barnes (ed.), *The Complete Works of Aristotle: The Revised Oxford Translation*, vol. 2. Princeton: Princeton University Press.
- Augustine (1993). *On Free Choice of the Will*, trans. T. Williams. Indianapolis: Hackett.
- Boethius (1887). *Commentarii in librum Aristotelis Peri hermeneias*, ed. C. Meiser. Leipzig: Teubner.
- Bonaventure (2005). *Disputed Questions on the Knowledge of Christ*, trans. Z. Hayes. Saint Bonaventure, NY: Franciscan Institute Publications.
- Bos, E. P. (ed.) (2013). *Medieval Supposition Theory Revisited*. Leiden: Brill.
- Brower, J. and S. Brower-Toland (2008). "Aquinas on Mental Representation: Concepts and Intentionality," *The Philosophical Review* 117(2): 193–243.
- Buridan, J. (2001). *Summulae de dialectica*, trans. G. Klima. New Haven: Yale University Press.
- Cesalli, L. (2007). *Le réalisme propositionnel, sémantique et ontologie des propositions chez Jean Duns Scot, Gauthier Burley, Richard Brinkley et Jean Wyclif*. Paris: Vrin.
- Cesalli, L. (2013). "Meaning and Truth," in A. Conti (ed.), *A Companion to Walter Burley: Late Medieval Logician and Metaphysician*. Leiden: Brill, 87–134.
- Conti, A. (2014). "Walter Burley," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), <http://plato.stanford.edu/archives/fall2014/entries/burley/>
- De Rijk, L. M. (1963–67). *Logica Modernorum: A Contribution to the History of Early Terminist Logic*, 2 vols. Assen: Van Gorcum.
- Duns Scotus, J. (1987). "Concerning Human Knowledge (from *Ordinatio* I, dist. 3, q. iv, art. IV)," trans. A. Wolter, in *Duns Scotus: Philosophical Writings*. Indianapolis: Hackett, 96–132.

⁷⁴ Supposition theory has been keenly studied in the twentieth century. For further references, see fn. 67, as well as Dutilh Novaes (2008) for a more thorough analysis of this theory in Ockham and others in the fourteenth century.

- Duns Scotus, J. (1994). "Six Questions on Individuation from His *Ordinatio* II. d. 3, part 1, qq. 1–6," trans. P. V. Spade, in *Five Texts on the Medieval Problem of Universals*. Indianapolis: Hackett, 57–113.
- Dutilh Novaes, C. (2007). *Formalizing Medieval Logical Theories: Suppositio, Consequentiae and Obligationes*. Dordrecht: Springer.
- Dutilh Novaes, C. (2008). "An Intensional Interpretation of Ockham's Theory of Supposition," *Journal of the History of Philosophy* 46(3): 365–94.
- Gilson, E. (1965). *The Philosophy of St. Bonaventure*, trans. I. Trethowan and F. J. Sheed. Patterson: St. Anthony Guild Press.
- Henry of Ghent (2008). *Henry of Ghent's Summa of Ordinary Questions. Article One On the Possibility of Human Knowledge*, trans. R. J. Teske. South Bend, IN: St. Augustine's Press.
- Klima, G. (2001). "Introduction," in *Summulae de dialectica*, trans. G. Klima. New Haven: Yale University Press.
- Kretzmann, N. (1970). "Medieval Logicians and the Meaning of the *Propositio*," *Journal of Philosophy* 67: 767–87.
- Lagerlund, H. (ed.) (2010). *Rethinking the History of Skepticism: The Missing Medieval Background*. Leiden: Brill.
- McCord Adams, M. (1982). "Universals in the Early Fourteenth Century," in N. Kretzmann, A. Kenny, and J. Pinborg (eds.), *The Cambridge History of Later Medieval Philosophy*. Cambridge: Cambridge University Press, 411–39.
- Nuchelmans, G. (1973). *Theories of the Proposition. Ancient and Mediaeval Conceptions of the Bearers of Truth and Falsity*. Amsterdam: North-Holland.
- Panaccio, C. (1992). "From Mental Word to Mental Language," *Philosophical Topics* 20: 125–47.
- Panaccio, C. (1999). *Le discours intérieur. De Platon à Guillaume d'Ockham*. Paris: Seuil.
- Pasnau, R. (1997). *Theories of Cognition in the Later Middle Ages*. Cambridge: Cambridge University Press.
- Pasnau, R. (2014). "Epistemology Idealized," *Mind* 122: 987–1021.
- Perler, D. (2000). "Essentialism and Direct Realism: Some Late Medieval Perspectives," *Topoi* 19: 111–22.
- Philip the Chancellor (1985). *Philippi Cancellarii Summa de bono*, ed. N. Wicki. Berne: Francke.
- Pouillon, H. (1939). "Le premier traité de propriétés transcendentales. La 'Summa de bono' du Chancelier Philippe," *Revue Néoscholastique de Philosophie* 42: 40–77.
- Spade, P. V. (1982). "The Semantics of Terms," in N. Kretzmann, A. Kenny, and J. Pinborg (eds.), *The Cambridge History of Later Medieval Philosophy*. Cambridge: Cambridge University Press, 188–96.
- Spade, P. V. (1994). *Five Texts on the Mediaeval Problem of Universals: Porphyry, Boethius, Abelard, Duns Scotus, Ockham*. Indianapolis: Hackett.
- Spruit, L. (1993). *Species Intelligibilis: From Perception to Knowledge, Classical Roots and Medieval Discussions 1*. Leiden: Brill.
- Stump, E. (2003). *Aquinas*. New York: Routledge.
- Tachau, K. (1988). *Vision and Certitude in the Age of Ockham: Optics, Epistemology and the Foundations of Semantics 1250–1345*. Leiden: Brill.
- Thomas Aquinas. (1882). *Quaestiones disputate de veritate [De veritate]*, in Leonine edition, *Opera Omnia*, Tomus XII, vol. 1, fasc. 2, QQ. 1–7; trans. R. W. Mulligan. 1952. *Truth, Questions 1–9*. Chicago: Henry Regnery.
- Thomas Aquinas. (2010). *Treatise on Human Nature: The Complete Text (Summa Theologiae I, Questions 75–102)*, trans. A. J. Freddoso. South Bend, IN: St. Augustine's Press.

- Van Dyke, C. (2009). "An Aristotelian Theory of Divine Illumination: Robert Grosseteste's Commentary on the *Posterior Analytics*," *British Journal for the History of Philosophy* 17: 685–704.
- Walter Burley. (1973). *Commentarius in librum Perihermeneias*, in S. F. Brown (ed.), "Walter Burley's Middle Commentary on Aristotle's *Perihermeneias*," *Franciscan Studies* 33: 45–134.
- William of Ockham. (1957). Prologue to the *Expositio super viii libros Physicorum*, in P. Boehner (ed. and trans.), *Ockham: Philosophical Writings*. Edinburgh: Thomas Nelson, 2–16.
- William of Ockham. (1974). *Ockham's Theory of Terms: Part I of the Summa Logicae*, trans. M. J. Loux. Notre Dame, IN: University of Notre Dame Press.
- William of Ockham. (1998). *William of Ockham: Ockham's Theory of Propositions, Part II of the Summa Logicae*, trans. A. J. Freddoso and H. Schuurman. South Bend, IN: St. Augustine's Press.

CHAPTER 3

EARLY MODERN THEORIES OF TRUTH

ALAN NELSON

IN early modern philosophy, as in almost all philosophy, the notion of truth is indispensable. In most of the period's canonical thinkers, however, a theory of truth emerges from other considerations that are taken to be more fundamental. Descartes's system of philosophy introduced a new focus on foundational epistemological considerations. In Descartes and those influenced by him, thus almost every important philosopher through Kant, problems about knowledge were addressed in terms of the states of individual minds. As philosophical thinking about truth became inextricably bound up with theories of mind, mental states themselves, usually called *ideas* or *concepts*, become either intrinsically true, true in relation to one another, or true in relation to extramental things. This new focus tended to have the effect of psychologizing (an anachronistic, but apt term) theories of truth.

The treatment of truth in early modern theories therefore contrasts with earlier and later theories that make use of abstract *propositions*. The word "proposition" is retained, but propositions tend to be regarded as constructions from concrete ideas in individual minds. One important consequence of this is that the natural assumption that linguistic structure is a guide to propositional structure is abandoned. This means that for the philosophers discussed in this chapter, sentences expressing common-sense facts like "The cat is on the mat" do not express propositions that share the sentence's linguistic form—not even under analysis. Thought is very decidedly prior to language in these theories.¹ A second important consequence of these psychologized theories is that there is a problem of the unity of composite ideas that closely parallels later discussions of the problem of the unity of propositions. This is especially true of propositions that do not have a simple subject-predicate structure, if indeed there are any without that structure allowed by these early modern theories. One virtue of theories that do employ

¹ For a general assessment of this point, see Buroker (1997). Hobbes is a qualified exception to the general point.

structured propositions is that they are subject to formal treatments of truth and inference. It should not be surprising then, that most (but not all) of the canonical early modern philosophers had anti-formalist, psychologistic conceptions of truth and logic.

We begin with an examination of Descartes who, more than anyone, set the agenda for the period in the middle of the seventeenth century. Spinoza is taken up next, for he pushed the Cartesian metaphysics of mind to something like a philosophical limit point. A section is devoted to Locke who in some ways works squarely within the Cartesian framework, but also introduces characteristic elements of empiricist theorizing. Locke is especially important in this regard for his treatment of truth, because more resolute empiricists such as Hobbes and Hume have hardly any use for a theory of truth.² We finally turn to Leibniz whose theory of truth incorporates or reacts to the theories of the others while making distinctive contributions of his own.

3.1 DESCARTES

A potentially frustrating feature of early modern philosophical theories is that truth is often taken to be comparatively unproblematic. Descartes expresses the attitude thus:

... for my part, I have never had any doubts about truth, because it seems a notion so transcendently clear that nobody can be ignorant of it ... What reason would we have for accepting anything which could teach us the nature of truth if we did not know that it was true, that is to say, if we did not know truth? (CSMK: 139)³

The hardest problems and most elaborate theoretical constructs arise instead in the analysis of falsity. Because most philosophers in this period held that a true *idea* cannot be obtained by performing some operation of negation on a false idea, questions about truth and falsity are not directly convertible. Another problem that faces us immediately when engaging Descartes's theories is his laxity with the word "truth." It should be no surprise that he and other early modern philosophers very often use "truth" in some of the non-technical senses found in ordinary dictionaries. Moreover, even in clearly theoretical contexts, some care is required to separate different notions of truth. In fact, it is sometimes difficult to say whether the same word is being used for different concepts or whether the concept itself is loose. The simplest notion of truth equates it with being. A good example of this occurs in the following pronouncements from a 1649 letter:

... there is no distinction between truth and the thing or substance that is true.

Truth consists in *being*, and falsehood only in *non-being*, so that the idea of the infinite, which includes all being, includes all that there is of truth in things, and

² In Hume's monumental, wide-ranging, and massive *A Treatise of Human Nature*, not a single paragraph is devoted to a theoretical notion of truth.

³ References to Descartes's writings are from the English translation in Cottingham et al. (1985–91). The three volumes of this work are abbreviated to CSM 1, CSM 2, and CSMK.

cannot contain anything false; and this is so even if you want to suppose it untrue that that infinite being exists.

And for good measure,

... for a being in which there is no imperfection cannot tend to non-being, that is, cannot have non-being, or non-good, or non-true as its end or purpose, since these three things are the same. (CSMK: 377–8)

Here we get a straight equation of truth and being. This makes existing things themselves into *truths* and there are no further questions to be asked about truths except for those regarding their proper inventory. Those further questions are explicitly metaphysical. This sense of “true” is not restricted to Descartes; Locke takes note of it more succinctly:⁴

Indeed, both *ideas* and words, *may* be said to be *true in a metaphysical sense* of the word Truth; as all other things, that any way exist, are said to be true; i.e. really to be such as they exist. (2.32.2)

Locke here has added the point that ideas and words are themselves true in this sense that he calls “metaphysical.” As we shall see, this can create confusion when we come to consider ideas and words as representing things. An idea of phlogiston falsely represents the world, but the idea exists as a state of particular minds and is, therefore, itself true in the metaphysical sense.

Both Descartes and Locke do move past the theoretically thin notion of metaphysical truth to another more familiar and more interesting notion. In the second passage from Descartes quoted above he, like Locke, characteristically effects a transition to speaking of ideas. Part of the point, of course, is to exploit a relation of conformity or correspondence between the “metaphysical truths” and the ideas that represent them. One might have thought this relation of correspondence should itself answer to the term “truth” and Descartes and Locke also find this appropriate. Descartes writes,

Of course, it is possible to explain the meaning of the word to someone who does not know the language, and tell them that the word “truth”, in the strict sense, denotes the conformity of thought with its object, but that when it is attributed to things outside thought, it means only that they can be the objects of true thoughts, either ours or God’s. (CSMK: 139)

Here Descartes seems to be saying that the metaphysical notion of truth as being is derivative. There is a primary, “stricter” notion that relates the mental to its “objects”—intentional objects as we would say today. He immediately goes on to muddle matters by

⁴ References to Locke’s *Essay Concerning Human Understanding* are cited thus: Book.Chapter. Section with original typography from Locke (1975 [1689]), ed. Nidditch. Locke’s own theory of truth is considered below.

referring to the mental items, the “thoughts,” as themselves *true*. A similar extreme terminological flexibility, not to say sloppiness, is also found in Locke:

Truth lies in so joining, or separating these Representatives, as the Things they stand for, do, in themselves, agree, or disagree: and *Falshood* in the contrary (2.32.19)

Locke is saying here that truth lies in the structure of mental representatives (i.e. ideas), but only insofar as this structure “stands for” or conforms to the structure of things, that is, metaphysical truths. Truth obtains when the “joining or separating” of ideas matches up with the ways in which properties do or do not belong to the things the ideas represent. Although Descartes and Locke might seem to suggest that this is the central concept of truth, the main focus of their analyses is on ideas themselves. Their characteristic terminological laxity then allows them to speak of truth as an intrinsic feature of the ideas that do conform to things. In conclusion, there are at least three senses of “true” at work here:

- (1) Anything that has being is (metaphysically) true.
- (2) A relation of (representative) truth can obtain between ideas and other things.
- (3) Some ideas are (intrinsically) true such that they represent objects truly; i.e. they yield truths.

Since the main theoretical work is done by (3), we need next to consider Descartes’s account of ideas more closely.

A good place to start is with Descartes’s famous dictum that intrinsically true ideas are *clear and distinct*. In the Synopsis of his *Meditations* he writes, “. . . everything that we clearly and distinctly understand is true in a way which corresponds exactly to our understanding of it” (CSM 2: 12).⁵ Descartes has three ways of explicating the special intrinsic character of clear and distinct ideas. The first is phenomenological—clear and distinct ideas are experienced “clearly” and “distinctly.”

I call a perception “clear” when it is present and accessible to the attentive mind—just as we say that we see something clearly when it is present to the eye’s gaze and stimulates it with a sufficient degree of strength and accessibility. I call a perception “distinct” if, as well as being clear, it is so sharply separated from all other perceptions that it contains within itself only what is clear. (CSM 1: 207–8)

Another phenomenological feature of clear and distinct ideas is that they are invariably affirmed; when an idea is clear and distinct it cannot be denied or even doubted.⁶

⁵ Since the perceptions of a mind, or the mind’s understanding always involve ideas, I shall use the modifier “clear and distinct” for both ideas and perceptions. For some discussion of the subtleties omitted here see Lennon (2008: ch. 5).

⁶ “Admittedly my nature is such that so long as I perceive something very clearly and distinctly I cannot but believe it to be true” (CSM 2: 48). For an explanation of this and additional texts, see Nelson (1997). As we shall see, Spinoza is highly critical of this second feature of Descartes’s true ideas.

The reason that clear and distinct ideas present this phenomenology is, of course, that human minds are specifically created that way by God. The second intrinsic feature of true ideas involves what Descartes calls their *objective reality*. He writes by way of definition, "By this I mean the being of the thing which is represented by an idea, in so far as this exists in the idea" (CSM 2: 113). One natural reading of this is that objective reality is the content of the idea in virtue of which it represents an object. Mental representation is a very tricky business, so it is no surprise that Descartes does not articulate a perspicuous general account of it. But in the case of a true idea, one's conscious awareness of the idea, and the special way in which the idea contains the being of the represented thing, constitutes knowledge of what the idea represents. This is again courtesy of the way the omnipotent being fashions thinking things.

Another way of explaining these features of true ideas emphasizes that they are innate to the mind. For Descartes, this is again tightly connected with the ultimate appeal to God's design of thinking things as having true ideas with the phenomenological and representative properties just noted. The innateness of true ideas is nevertheless a separate consideration. Other philosophers might take the innateness of true ideas to be a brute fact; as we will see, Spinoza was close to holding that. Still others might give an evolutionary biological account of how the mind comes to have certain structures that are true of the world. And still others like Locke will mostly agree with Descartes on the phenomenology and partly agree with him on representation, but reject innate ideas.

For Descartes, innate ideas must be true ideas because their presence in the mind is due to the omnipotence and infinite benevolence of the creator. By contrast, ideas that are not innate are acquired by sensing the world or by the mind's constructing them from sensory materials. These processes are obviously fallible, unlike the "process" by which God produces minds in the first place. On the one hand, Descartes states, "there can be no falsity save in composite [ideas] which are put together by the intellect" (CSM 1: 32). And on the other hand he asks, "The more skilled the craftsman the more perfect the work produced by him; if this is so, how can anything produced by the supreme creator of all things not be complete and perfect in all respects?" (CSM 2: 38). Here we have our second way of dividing true ideas from false ideas. The true ideas have an ontological ground in the mental structure created by a non-deceiving God while the false ones involve compositions produced by the finite mind's fallible judgments. And this gives us another perspective on the phenomenology of true ideas. In virtue of the way it is created, the mind's nature is such that the true ideas are clear, distinct, and automatically affirmed. This train of claims about innateness has struck Descartes's critics as highly problematic.

Hobbes, for example, pointed out to Descartes in his *Objections* published along with Descartes's *Meditations* and the latter's replies that those in a dreamless sleep do not experience innate ideas (CSM 2: 132). And the entire first Book of Locke's *Essay* is a polemic against innate ideas. A recurring argument that Locke uses there against nativism is that very young children are not aware of them (e.g. 1.4.13). These empiricist thinkers require that all ideas be furnished by experience. A possible response would be that innate ideas are indeed always experienced, but they are experienced when then are mixed up with other, false ideas. In other words, they are not always clear and distinct, but are

“confused.” Descartes uses the term “confused,” literally con-fused, or fused together, for the false ideas that result from improper composition. Again, “there can be no falsity save in composite [ideas] which are put together by the intellect.” Descartes can say, therefore, that all minds have the disposition to clearly and distinctly perceive innate ideas. This need not be a mysterious, unanalyzed disposition; the innate ideas in confused composite ideas can be made clear and distinct by distinguishing them from the confusing components: “For I shall unquestionably reach the truth, if only I give sufficient attention to all the things which I perfectly understand, and separate these from all the other cases where my apprehension is more confused and obscure” (CSM: 243). And this distinguishing is to be accomplished by the process of methodical mediation that Descartes prescribes. Of course, these considerations will gain no traction with those who do not come to enjoy the promised benefits of Cartesian meditation.

The third, closely related way that clear and distinct, true ideas are special is that they are purely intellectual; they are *distinct* from the characteristic contribution of sensation. In the strictest sense Descartes’s true ideas are, therefore, true representations of only God, and of the essences of mind and body—thought and extension, that is, and perhaps carefully constructed determinations of these and abstractions from them. So a cube for example is a determination of extension, and the idea of finite substance is abstracted from the ideas of mind and body. In Descartes’s case, his ontologically parsimonious substance dualism with respect to metaphysical truths means that the set of corresponding representative truths is sharply restricted. The perceptions expressed by ordinary empirical propositions such as “Grass is green” turn out not to be true in the strictest sense. The idea of grass cannot be made clear and distinct even though the generic essence of any sample of grass, namely extension, can be. Lacking a clear and distinct perception, there is no guarantee of truth—even in cases that we might habitually take to be absolutely true.⁷

Descartes’s framework was subjected to both metaphysical and epistemological criticism. The former is emphasized by Spinoza, who significantly modifies Descartes’s ontology while relying even more on true ideas. Epistemological difficulties with the representative features of both Descartes’s and Spinoza’s true ideas are stressed by Locke.

3.2 SPINOZA

In Spinoza’s mature philosophy, the notion of truth is invested with proprietary metaphysics even more heavily than in Descartes’s. Here is a very brief overview of the relevant doctrines. Spinoza is a substance monist, which means that for him there is only one thing that has independent existence. Spinoza calls the one substance “God.” Anything else has to be “in” God—a mode of God—literally a *way* in which God exists. Moreover, God is

⁷ For further discussion of some of the complications raised by the last two paragraphs, see Nelson (2008).

both a thinking substance and an extended substance. This means that finite modes of God including such familiar objects as human beings, plants, and rocks also have both mental and physical aspects. Spinoza unites these dual or parallel aspects in the human being by saying that the human mind *is* the idea of the human body. That states both an identity and that “The object of the idea constituting the human Mind is the body, or a certain mode of extension that actually exists and nothing else” (E2P13).⁸ But this result holds for all finite modes,

For the things we have shown so far are completely general and do not pertain more to man than to other Individuals, all of which, though in different degrees, are nevertheless animate . . . And so, whatever we have said of the idea of the human Body must also be said of the idea of any thing. (E2P13Corollary)

Now we come to true ideas. Spinoza writes, “A true idea must agree with its object” (E1Axiom6). This crucial axiom appears to express the representational feature of true ideas that is found in Descartes, but the appearance is deceiving. Since E2P13 and its corollary mean that ideas *are* their objects, the “agreement” mentioned in the quoted axiom is very tight indeed. It can be thought of as an identity under dual “aspects” that Spinoza calls *attributes*. Obviously no idea that a human being consciously experiences is ever true in this sense. The true ideas are ideas of God’s: “All ideas, insofar as they are related to God, are true” (E2P32). The true ideas constitute God’s infinite intellect, but they also *are* the finite minds which, it must be remembered, are the ideas of the bodies.

Human minds cannot *have* true ideas of bodies; human minds *are* true ideas of God’s. In other words, a true idea of a body cannot be a human mental state—it is not true when “related to the human mind.” Spinoza does allow, however, that humans can have ideas that are *adequate*:

By adequate idea I understand an idea which, insofar as it is considered in itself, without relation to an object, has all the properties or intrinsic denominations of a true idea. (E2Definition4)

This is a difficult definition to work with because Spinoza is not entirely clear about what the intrinsic properties of ideas are supposed to be. It is clear, however, that he is here regarding an idea’s object to be extrinsic to it. Spinoza’s adequate ideas, then, are perhaps closer to what one would ordinarily take to be relevant to the most common philosophical concept of *truth*. This is borne out by the fact that ideas cannot fail to be true, but they can fail to be adequate and it is the latter that Spinoza regards as *false* ideas: “Falsity consists in the privation of knowledge which inadequate, or mutilated and confused, ideas involve” (E2P35). It is interesting that Spinoza here explicitly connects falsity with imperfect knowledge. And this he does in two ways.

⁸ References to Spinoza’s *Ethics* are from the English translation in Curley (1985). The abbreviation E2P13 e.g. is read: Ethics, Part 2, Proposition 13.

The first is that false ideas are “mutilated.” This colorful term indicates that when a human mind *has* an inadequate idea, there is a failure of agreement between that idea and its object. All ideas are true, and therefore agree with their object “when referred to God,” so inadequate ideas fail to agree when referred to the human because the human misidentifies the object.⁹ When someone has an idea of a tree in front of her, the object of that idea referred to God is a part of the human body; today we might think of it as a state of the body’s system of visual perception. But the non-agreeing object of the idea referred not to God, but to the mind that has the idea, is the tree. Of course the tree causally affects the mind, but to take the tree as the object of the idea is to “mutilate” it. The human mind inadequately perceives the object of the idea that is true when referred to God only “partially” (E2P11Corollary).

This introduces the second way in which falsity connects with imperfect knowledge. When the human mind has an inadequate idea it perceives its object only “partially” because that object is mixed up with or “confused” with an object external to the idea. This is one reason that E2P35, quoted above, says of the inadequate and mutilated ideas that they are also confused. This strikingly recalls what we found in Descartes, where confusion is the opposite of distinctness. And distinctness, along with clarity, is a phenomenological feature of ideas that are true (in Descartes’s sense) and therefore constitute knowledge. So in Spinoza, we have it that an idea that is consciously perceived in a confused way (despite its being true in God’s intellect) misperceives its object and is, therefore, a false idea. This characterization of inadequate ideas now points the way to the positive notion of adequate ideas.

Spinoza’s formal definition of adequate ideas requires that they have all the properties of true ideas once the consideration of the object is set aside. The most important of these intrinsic properties is clarity and distinctness. And as in Descartes, clear and distinct ideas constitute knowledge even though, to repeat, Spinoza regards these ideas as “adequate” rather than “true.” This raises a puzzle about how it is that adequate ideas can constitute human knowledge, given that they cannot be considered true. The general outline of Spinoza’s solution is plainly indicated, although the details are left murky. When a human mind clearly and distinctly perceives what is common to all objects, that idea will be adequate. In E2P38Corollary Spinoza explains that, “. . . all bodies agree in certain things which must be perceived adequately or clearly and distinctly by all.” He is thinking that all bodies are in some state of motion and rest and, ultimately, that they are all extended. The solution seems to rely on the fact that when a human mind clearly and distinctly perceives an adequate idea it achieves the perspective of God’s infinite intellect without, needless to say, having the entire universe as an object agreeing with the idea. Only God’s intellect can literally agree with (i.e. be one and the same as) the entire universe of bodies. In this way, adequate ideas serve as the human mind’s surrogates for true ideas. There is one further twist. Spinoza holds that any true idea agrees with its object, which is a body. But every body, and not just every human body, is one and the

⁹ This is made explicit in E2P36Demonstration.

same thing as an idea. From this it follows that there is an idea of every idea. Finally, if a human mind has a clear and distinct, adequate idea, it will also have the idea of that idea. This second order idea will agree with its object (i.e. the idea of which it is an idea). And this finally means that the second order idea is a true idea in the strict sense of Spinoza's definition. So the human mind can, after all, have some true ideas.¹⁰

Spinoza mobilizes his analysis of true ideas to criticize what he took to be some faulty doctrines of Descartes'. Descartes famously suggests in the "Fourth Meditation" that a true *judgment* results from the joint operation of the intellect and the will. When an idea that is clear and distinct in the intellect is affirmed by the will, all is well, but when the will affirms a confused idea an erroneous judgment is produced. Spinoza's extreme intellectualism blocks a deep distinction between intellect and will. This is partly driven by his determinism, but there is also a connection with truth. The dual faculty theory of judgment seems to make clarity and distinctness an extrinsic property of true ideas. It is as if one seeking the truth must check a candidate idea to see whether it displays the property of clarity and distinctness. If it does, then it is proper for the will to affirm the idea; otherwise the will should be left in suspense until the confusion is removed. In this way, Cartesian clarity and distinctness takes on the aspect of a criterion of truth that can be wielded against skeptical worries. Against this, Spinoza writes that "truth is its own standard" because one cannot know that one knows something unless he first knows it.¹¹ This in turn reinforces Spinoza's doctrine that a true idea is itself an active knowing and does not stand in need of affirmation from an external faculty of will.¹²

3.3 LOCKE

In Descartes, and especially in Spinoza, theorizing about truth focuses on the unitary true idea. True ideas are true because of their connection with their objects, but the nature of this connection is taken to be axiomatic or a brute phenomenological given. A move away from this emphasis is evident in the theories of philosophers like Hobbes and Locke who tend more toward more epistemically modest, empiricist lines of thought. This alternative to unitary true ideas is noticed by Descartes. It is succinctly expressed in the following quote, which in its original context, appears as a formal definition.

When we say that something is *contained in the nature or concept* of a thing, this is the same as saying that it is true of that thing, or that it can be asserted of that thing. (CSM 2: 114)

¹⁰ Any interpretation of Spinoza on these points will involve some speculation. For a fuller discussion of some of the difficulties that are involved see Della Rocca (1996: ch. 6).

¹¹ This is found in the Scholium to E2P43 which itself reads, "He who has a true idea at the same time knows that he has a true idea, and cannot doubt the truth of the thing."

¹² It might be that Spinoza and many following him have misread Descartes as holding that clarity and distinctness is an extrinsic property of true ideas. For discussion see Lennon (2008: ch. 5).

Here Descartes is defining the italicized phrase, but it also clearly provides a definition of the formula:

P is true of *S*

where *P* is contained in the concept of *S*, and *S* is a “thing.” Descartes makes very limited use of this *concept containment* notion of truth, but it foreshadows the more explicit treatments we shall examine next. Part of its significance is that it suggests that truth involves propositional structure. As stated in Descartes’s definition it is a simple subject-predicate structure that mirrors the substance-mode ontology that is common to nearly all early modern philosophers. Here is the way in which Locke associates truth with propositions:

Truth then seems to me, in the proper import of the Word, to signify nothing but *the joining or separating of Signs, as the Things signified by them, do agree or disagree with one another*. The *joining or separating* of signs here meant is what by another name, we call Proposition. So that Truth properly belongs only to Propositions: whereof there are two sorts, *viz.* Mental and Verbal; as there are two sorts of Signs commonly made use of, *viz.* *Ideas* and Words. (4.5.3)

Locke acknowledges that there is also a use of “true” that applies to ideas in the more Spinozistic sense, but rejects it as an analysis of truth:

Though Truth and Falsehood belong, in Propriety of Speech, only to Propositions; yet *Ideas* are often times termed *true or false* . . . Though, I think, that when *Ideas* themselves are termed true or false, there is still some secret or tacit Proposition, which is the Foundation of that Denomination: as we shall see, if we examine the particular Occasions, wherein they come to be called true or false. (2.32.1)¹³

One notable feature of Locke’s main definition of truth at 4.5.3 is that he clearly marks its application to both ideas and words. We also find this in Hobbes. In his *Leviathan* he writes,

When two names are joined together into a consequence or affirmation (as thus, *a man is a living creature*, or thus, *if he be a man, he is a living creature*), if the latter name, *living creature*, signify all that the former name, *man*, signifieth, then the affirmation or consequence is *true*, otherwise *false*. For *true* and *false* are attributes of speech, not of things. And where speech is not, there is neither *truth* nor *falsehood*. (1.4.11)¹⁴

¹³ This focus on propositional structure tends to align with the epistemic modesty found in philosophers typically regarded as empiricists, but this is only a tendency. The arch-empiricist Hume regarded it as “a very remarkable error” (Hume 2000 [1738–40]: 67) to separate the bare conception of ideas from propositions. But Hume also has virtually no use for a theoretical treatment of truth.

¹⁴ References to *Leviathan* are cited thus: Part.Chapter.Paragraph from Hobbes (1994 [1651]), ed. Curley.

Hobbes is unusual in this period for making truth apply primarily to speech rather than to ideas or “things.” Although Locke makes the distinction between mental and verbal propositions, he plainly regards truth as primarily a matter of mental propositions and only secondarily a matter of verbal propositions. To understand Locke’s theory, we must now examine what he means by the “joining and separating” of ideas that generate propositions. It is clear that “joining” in both Hobbes and Locke involves affirmation and “separation” involves denial. The standard view was that the joining and separating of words proceeds according to conventions that make words into appropriate signs for ideas.¹⁵ But in what conditions are ideas rightly joined or separated to form true mental propositions? Hobbes may well be committed to saying that this is also a matter of convention. Most early modern philosophers would have regarded that as a profoundly skeptical, abhorrent position. Locke does not answer the question of how the joining and separating of ideas works in the context of truth itself. In keeping with the early modern character of his theory, Locke engages epistemology to articulate the theory of truth. The passage from 4.5.3 quoted above indicates that Locke held that knowledge of the world requires having ideas joined and separated as things in the world agree and disagree, that is, as things in the world are interrelated. Locke explicitly defines knowledge this way:

Since *the Mind*, in all its Thoughts and Reasonings, hath no other immediate Object but Its own *Ideas*, which it alone does or can contemplate, it is evident, that our Knowledge is only conversant about them. (4.1.1)

Knowledge then seems to me to be nothing but *the perception of the connexion and agreement, or disagreement and repugnancy of any of our Ideas*. In this alone it consists. (4.1.2)

This shifts the question about the joining and separating of ideas to a question about *perceiving* agreements and disagreements between ideas. Although Locke is not as explicit as one would wish on the crucial question, a coherent account does emerge from the text of the *Essay*. To perceive that idea A agrees with idea B is to perceive that idea B is contained in idea A as a component of it; Locke holds a version of the concept containment theory of truth as articulated in the passage from Descartes quoted above. This idea containment theory means that to perceive agreement is to perceive identity—B as contained in A is identical with B occurring separately.¹⁶

This theory, like Descartes’ and Spinoza’s, sharply restricts the truths we can know. Unless the perception of agreement is clear and distinct (which Locke glosses as

¹⁵ Hobbes stood apart in making propositional thought depend on words. His view is sharply criticized in the extremely influential *Port-Royal Logic*. See e.g. Arnauld and Nicole, trans. Buroker (1996: 27–8).

¹⁶ Also, to perceive disagreement is to perceive non-identity; for convenience, I shall leave that complementary point tacit in what follows. Also, I am here switching from “concept” to “idea” to conform with Locke’s usage. Early modern philosophers who use the term “concept” almost always use it synonymously with “idea.” In attributing this theory of truth to Locke, I am following the convincing account in Newman (2007: 333–42).

intuitive or as demonstrated from intuitions)¹⁷ there can be no knowledge, but only probable opinion. This might seem implausible in the case of highly reasonable opinions, so Locke introduces a second device for handling these cases:

Thus the Mind has two Faculties, conversant about truth and Falshood. First, *Knowledge*, whereby it certainly perceives, and is undoubtedly satisfied of the Agreement or Disagreement of any *Ideas*. Secondly, *Judgment*, which is the putting *Ideas* together or separating them from one another in the Mind, when their certain Agreement or Disagreement is not perceived, but *presumed* to be so; which is, as the Word imports, taken to be so before it certainly appears. And if it so unites, or separates them, as in Reality Things are, it is *right Judgment*. (4.14.4).

The key here is that “judgment” issues in opinion, not knowledge, when ideas are *presumed* to agree. When one takes oneself to have strong reasons, the phenomenology of presumed agreement might be very similar to that of perceived agreement, but only the latter issues in knowledge. If, however, ideas that are merely presumed to agree do in fact agree, that is if the proposition is indeed true, we instead call the result “right judgment.” In the theories of Descartes and Spinoza, the restriction on knowable truths is offset by the robustness of the truths we do know. One might worry, however, that the way Locke psychologizes truth threatens to sever any connection between ideas agreeing in thought and things agreeing outside the mind. How are true mental propositions related to how “in Reality Things are?”

Locke addresses this aspect of truth using another feature of his general theory of ideas. Ideas can be either simple or complex. Complex ideas are formed from other ideas by mental operations (2.12.1). Simple ideas are passively received by the mind through either sensation or by “the notice the mind takes of its own operations,” which Locke calls reflection (2.1.1–4). It is simple ideas, therefore, that ultimately secure the mind–world connection.

Because being nothing but the effects of certain Powers in Things, fitted and ordained by GOD, to produce such Sensations in us, they cannot but be correspondent, and adequate to those Powers: And we are sure they agree to the reality of Things. (2.31.2)

The connection is tenuous because Locke’s simple ideas, by their very nature, have severely limited content. Each contains “nothing but one uniform appearance” (2.1.1). This is the cost Locke must accept for the benefit of doing without innate ideas.

This last feature of Locke’s theory of truth and knowledge fits with his interesting view of empirical scientific propositions. The “perceived agreement” that constitutes knowledge of a proposition requires either direct clear and distinct perception (i.e. “intuition”) or else clear and distinct demonstration from intuitions. This means that propositions

¹⁷ 4.3.2. I am omitting discussion of Locke’s very difficult treatment of what he calls “sensitive knowledge.” This case is shown to be assimilated to the perception of agreement in Priselac (2016: ch. 2).

of empirical science, no matter how strongly supported by reason, can be only judged and not known. There could, however, be knowledge of a moral science that was, like mathematics, generated by demonstration from intuited definitions (section 3.11.16). It is natural to compare Locke and Hobbes on this point. Hobbes was reaching for a more formal theory of concept containment. He characterized true propositions as expressing a kind of numerical equality. Inference or “ratiocination” was then understood to be a kind of calculation of sums and differences (1.4.14, 1.5.1). Scientific knowledge results from correctly calculating conclusions from precise definitions that are true in virtue of expressing equalities. This picture bristles with difficulties and Hobbes does not provide enough detail to even speculate about how he might have carried out the project.¹⁸ Nevertheless, the ideal of formalizing truth and inference in a way consistent with the early modern emphasis on the mental proved extremely influential. It certainly influenced Leibniz’s theory of truth, to which we now turn.

3.4 LEIBNIZ

In Leibniz we find the most highly developed theory of truth in the early modern period. The innovations and elaborations he introduces are, nevertheless, best understood as developments, albeit sometimes radical developments, of the theories discussed above. Concept containment is the now familiar, central theoretical notion employed by Leibniz. A good deal of what is distinctively Leibnizian in his theory of truth can be seen as deriving from his extraordinary theory of concepts and the way in which they connect up with metaphysical and epistemological concerns.

Leibniz’s treatment of such familiar concepts as “chair” or “triangle” depends on the more fundamental treatment of the concepts of substances. As is well known, Leibniz’s metaphysics requires that there exists an infinity of individual finite substances which he eventually calls *monads*. Monads are unextended thinking things that perceive other monads. For our purposes, the most important thing about monads is that each has a concept that is *complete* in a very strong sense of that term. This explains why triangles are not substances. They come in different sizes and shapes, for example, so their concepts are not fully specific and are, therefore, incomplete. The complete concept of an individual is infinitely complex and it includes,

That each singular substance expresses the whole universe in its own way, and that all its events, together with all their circumstances and the whole sequence of external things are included in its notion. (AG41)¹⁹

¹⁸ For some additional discussion, see Pettit (2008: ch. 3).

¹⁹ References to Leibniz’s writings are from the English translation by Ariew and Garber (Leibniz 1989), which is abbreviated to “AG.”

When this extraordinary idea is combined with Descartes's simple definition quoted above, which links truth and concept containment, we get what can be called the *complete concept theory of truth*, or the *predicate in subject theory of truth*:

It is common to every true affirmative proposition, universal and particular, necessary or contingent, that the predicate is in the subject, that is, that the notion of the predicate is somehow contained in the notion of the subject. (AG95)

There is nothing that can be truthfully affirmed of a substance that is not already contained in its concept. By construction then, every truth about a substance holds in virtue of what is intrinsic to it. This might seem illegitimately to conflate the substance itself with its concept. We have not yet said what the "complete concept of a thing" is supposed to be; what is its metaphysical status? It would be very difficult to maintain that these infinitely complex concepts reside in finite minds. Things become clearer, however, when it is recognized that an infinite mind would be well suited to conceive these concepts. The omniscient, infinite intellect knows them perfectly despite their infinite complexity and actualizes the complete concepts by creating them. The comparison with Spinoza on this point is interesting. As we have seen, Spinoza maintains that finite things *are* ideas that compose the infinite intellect, so he says that they "follow" from the infinite intellect. Leibniz's God, by contrast, creates *ex nihilo* substances that perfectly match his ideas of them.

It is a striking consequence of Leibniz's theory of truth that every true proposition about a substance is analytic in the now familiar sense that a concept is identically present in the subject and predicate of a true proposition. He is more explicit about this than Locke and has a much more fully articulated theory than Hobbes. To see this consider the proposition "All right triangles are triangles." A natural explanation of its truth is that the concept "right triangle" literally includes the concept "triangle." The agreement (to use Locke's term) between subject and predicate concepts consists in the concept "triangle" appearing in each. Leibniz indeed puts this by saying that every true proposition can be reduced to an "identity" (because he is happy to call propositions of the form " $AB = A$ " identities). In the geometrical example being considered the identity is obvious. It is less obvious that the concept "right triangle" contains the concept "the square of the adjacent sides is equal to the square of the hypotenuse," but this can be demonstrated to be an identity (again in Leibniz's special sense) with relative ease. Now consider a true proposition about a substance: "Caesar crossed the Rubicon." This, too, is analytic insofar as it is an identity statement because "Rubicon-crossing" is included in Caesar's concept just as "triangle" is included in "right triangle."

Although Leibniz could not have objected to the charge that his theory of truth makes all truths analytic, he is concerned to preserve a distinction between necessary and contingent truths. So although identities can be demonstrated from both the proposition about triangles and the proposition about Caesar, Leibniz does not want them to demonstrably follow in the same way. True propositions about geometrical subjects

follow with the kind of absolute necessity that almost everyone in the period would regard as unacceptably antithetical to human freedom. The only clear exception among canonical philosophers was Spinoza who embraced the conclusion that all of nature, including humanity, is determined.²⁰

Leibniz has two interconnected solutions to this problem about how to distinguish contingent truth from necessary truth. The first is that truths about substances follow from their concepts only on the *hypothesis* that God creates them. This means that they are not “absolutely necessary” as they are in Spinoza’s theory; they are merely *hypothetically necessary*. The reason that the creation of Caesar can be called a hypothesis is that the contrary is possible: other worlds not containing Caesar are possible because God understands them as alternatives for creation without creating them. This means that God, anyway, has some kind of choice about which world to create, so God is not subject to an absolute necessity. This in turn yields a sense, admittedly a thin sense, in which Caesar’s crossing the Rubicon necessarily follows from his concept without its being absolutely necessary.

The second solution invokes the notion of infinite analysis.

But in contingent truths, even though the predicate is in the subject, this can never be demonstrated, nor can a proposition ever be reduced to an equality or to an identity, but the resolution proceeds to infinity, God alone seeing, not the end of the resolution, of course, which does not exist, but the connection of the terms or the containment of the predicate in the subject, since he sees whatever is in the series. (AG96)

Our finite minds can access absolutely necessary truths as necessary because, unlike contingent truths, they do not require infinite intellects to infinitely analyze or know a priori. On the contrary, they are shaped to finite intellects insofar as they are finitely analyzable into identities. This might seem to make the distinction between necessary and contingent truths into an epistemological distinction, because contingent truths are “really” necessary; it is just that we cannot see the necessity as an omniscient being can. The two solutions are, nevertheless, connected in Leibniz’s philosophy. Since a complete concept expresses, “however confusedly, everything that happens in the universe” (AG42), the universe or world of infinitely many actually created monads is perfectly interconnected—this is Leibniz’s famous *pre-established harmony*. Without there being any causal interaction among the monads, their perceptions evolve in a maximally coherent way. The unanalyzable complexity of any actual complete concept is therefore bound up with its being actualized along with the rest of the universe. Now, given Leibniz’s rejection of trans-world identity (Mason, ed. 1967: 76), no other possible worlds contain Caesar. They might contain Caesar counterparts, but not the Caesar of

²⁰ “It is of the nature of reason to regard things as necessary, not as contingent” (E2P44).

our, actual, world. From this, we have that what is contained in Caesar's complete concept follows by a hypothetical rather than an absolute necessity.

I have been trying to explain Leibniz's theory of truth by describing how it is connected with some of his other distinctive doctrines, especially the complete concept theory. But it is probably not right to say that the theory of truth is derived from those other doctrines.²¹ Indeed, Leibniz often writes as though the theory of complete concepts is derived from the theory of truth:

Now it is evident that all true predication has some basis in the nature of things and that, when a proposition is not an identity, that is when the predicate is not explicitly contained in the subject, it must be contained in it virtually . . . Thus the subject term must always contain the predicate term, so that one who understands perfectly the notion of the subject would also know that the predicate belongs to it. Since this is so, we can say that the nature of an individual substance or of a complete being is to have a concept so complete that it is sufficient to contain and to allow us to deduce from it all the predicates of the subject to which this notion is attributed. (AG41)

Texts like this one led Bertrand Russell (1937 [1900]) to conclude that Leibniz derived most of the signature doctrines that we have noted from his subject-predicate logic, that is his theory of truth. What then is the ultimate ground of the theory of truth? In contexts such as the one just quoted, Leibniz is taking this conception of truth to be philosophically obvious; he did not expect resistance on the theory of truth, but only on the consequences that he drew from it. Referring to his use of the *predicate in subject* principle, Leibniz writes,

Finally I have given a decisive argument which in my view has the force of a demonstration; that always, in every true affirmative proposition, necessary or contingent, universal or particular, the concept of the predicate is in a sense included in that of the subject; the predicate is present in the subject; *or else I do not know what truth is*. (Mason, ed. 1967: 63)

This clearly echoes Descartes's remark quoted above that truth is something so "transcendently clear that no one can be ignorant of it." Once one sees that truth simply is concept containment, this serves as a fundamental *premise* for "decisive arguments" in metaphysics. It is not something that itself requires demonstration from other prior premises. Leibniz's philosophy is so thoroughly systematic, however, that any important doctrine in it can be seen to follow from the others. For our purposes it thus seems appropriate to begin with Leibniz's commitment to complete concepts for individual substances and to see how it issues in his theory of truth.

²¹ For more on this, see Sleight (1982).

3.5 CONCLUSION

Despite their richness, early modern theories of truth run into something of a philosophical dead end. Despite a vague vision in Hobbes and some halting advances by Leibniz, truth is mostly detached from formal considerations because of the concept containment principle. This means that the mind–world connection that a theory of truth is supposed to articulate is left mostly unarticulated. The connection depends on what is ultimately a conceptual or psychological basis, which is then paired with a dogmatic appeal to infinite being. And the only sort of conceptual relations straightforwardly dealt with resemble conjunction, perhaps supplemented by a the possibility of some kind of mental synthesis of the “conjuncts.” This is true even in Leibniz, so there is some justice to Russell’s charge that Leibniz and his early modern predecessors were hamstrung by their lack of an appropriate formal logic (1937 [1900]: xii). Theorizing primarily based on concept containment was not significantly challenged before Kant’s critical philosophy. In the metaphysical deduction of the categories in his *Critique of Pure Reason* Kant proposed that thought is propositionally structured by functions that involve more than neo-Cartesian ideas and more propositional forms than just predication. Today, the details of Kant’s own theory are themselves mostly of historical interest. But the correlation between thought and abstract propositional forms he pioneered makes possible the rich formal treatments of truth that we find in contemporary philosophy.

REFERENCES

- Arnauld, A. and P. Nicole (1996). *Logic or the Art of Thinking*, trans. and ed. J. V. Buroker. Cambridge: Cambridge University Press.
- Buroker, J. V. (1997). “The Priority of Thought to Language in Cartesian Philosophy,” in L. Falkenstein and P. Easton (eds.), *Logic and the Workings of the Mind*. Atascadero, CA: Ridgeview, 97–108.
- Cottingham, J., R. Stoothof, D. Murdoch, and (vol. 3 only) A. Kenny (trans. and eds.) (1985–91). *The Philosophical Writings of Descartes* (3 vols.), Cambridge: Cambridge University Press [CSM/K].
- Curley, E. (trans. and ed.) (1985). *The Collected Works of Spinoza*, vol. 1. Princeton: Princeton University Press.
- Della Rocca, M. (1996). *Representation and the Mind-Body Problem in Spinoza*. Oxford: Oxford University Press.
- Hobbes, T. (1994) [1651], *Leviathan*, ed. E. Curley. Indianapolis: Hackett.
- Hume, D. (2000) [1738–40]. *A Treatise of Human Nature*, ed. D. Norton and M. Norton. Oxford: Oxford University Press.
- Leibniz, G. W. (1967). *Philosophical Essays*, trans. and ed. R. Ariew and D. Garber. Indianapolis: Hackett. [AG]
- Lennon, T. (2008). *The Plain Truth*. Leiden: Brill.
- Locke, J. (1975) [1689]. *An Essay concerning Human Understanding*, ed. P. Nidditch. Oxford: Oxford University Press.

- Mason, H. (trans. and ed.) (1967). *The Leibniz–Arnauld Correspondence*. Manchester: Manchester University Press.
- Nelson, A. (1997). “Descartes’ Ontology of Thought,” *Topoi* 16: 163–78.
- Nelson, A. (2008). “Cartesian Innateness” in J. Broughton, and J. Carriero (eds.), *A Companion to Descartes*. Oxford: Blackwell, 319–33.
- Newman, L. (2007). “Locke on Knowledge,” in L. Newman (ed.), *The Cambridge Companion to Locke’s Essay Concerning Human Understanding*. Cambridge: Cambridge University Press, 313–15.
- Pettit, P. (2008). *Made with Words: Hobbes on Language, Mind, and Politics*. Princeton: Princeton University Press.
- Priselac, M. (2016). *Locke’s Science of Knowledge*. New York: Routledge.
- Russell, B. (1937) [1900]. *A Critical Exposition of the Philosophy of Leibniz*. London: George Allen & Unwin.
- Sleigh, R. (1982). “Truth and Sufficient Reason in the Philosophy of Leibniz,” in M. Hooker (ed.), *Leibniz: Critical and Interpretive Essays*. Minneapolis: University of Minnesota Press, 209–42.

CHAPTER 4

IDEALISM AND THE QUESTION OF TRUTH

CLINTON TOLLEY

4.1 INTRODUCTION: TRUTH AFTER THE COPERNICAN TURN

ON a fairly traditional—and, perhaps, fairly “platitudinous”¹—understanding of its nature, truth seems to be something that should merit the label of being “Janus-faced.” This is because truth appears to be something that points at the same time, and seemingly with equal necessity, toward both objects and subjects. On the one hand, the notion of truth seems to involve the notion of objectivity. If something is true, then (aside, perhaps, from certain cases e.g. involving self-reference) it is true whether we (subjects) like it or not, regardless of whether we think (believe, take) it to be true. What *makes* things true are simply the facts of the matter. On the other hand, the notion of truth also seems to involve an intrinsic connection to the notion of subjectivity—more specifically to the mental and linguistic activity of thinking, speaking subjects. This connection shows itself in two ways: first, in the fact that the truth sets a standard of *correctness* for us, is something of value to us, something the possession of which seems to form the end or aim of belief and the expression of which seems to form the goal of assertion; second, in the fact that what have seemed to many to be the most natural candidates for the things which are true—the things that *bear* the property of being true—are *our* representations of things, whether mental or linguistic, things, that is, that seem only to make sense to talk about in the context of representers.

This traditional conception is more or less encapsulated in what, for a long while, has served as the core definition of truth: “truth is the agreement of our thoughts with

¹ This is Crispin Wright’s description of “very general, very intuitive principles” such as “that to assert is to present as true” and “that to be true is to correspond to the facts” (Wright 1992: 34).

their objects”—“*veritas est adaequatio rei et intellectus*,” as Aquinas puts it.² To be sure, not everyone has accepted that all three of these aspects (relation to objective facts (*res*), standard for correctness (*adaequatio*), and relation to acts of representing by subjects (*intellectus*)) must be incorporated into the analysis of truth. Even so, philosophical accounts of truth that leave no room for, or cannot do justice to, one or another of these aspects are often, and for this reason, viewed with a certain amount of suspicion.

Almost immediately after it burst on the scene, Kant’s idealism came in for just such a criticism. In particular, there is often concern that Kant’s views will eliminate a place for the first two features of truth identified above. In Kant’s hands, idealism consists in the belief in the mind- or representation-*dependence* of certain aspects of the sensible world. But then, to the extent that what allegedly makes our representations of these aspects true is now a function of these representations themselves, idealism seems to imply, first, that our representing something to be so *is* itself responsible for its being so, and secondly, and correlatively, that there is no possibility for us to be mistaken in our representation of these aspects, since there is no further way that these aspects are “in themselves” above and beyond how we represent them as being, such that we could represent them falsely, or in a way in which they, in fact, were not.

Now, as we will see below (section 4.2), Kant himself did not think that his views required giving up on the traditional definition of truth. Yet as will also emerge, many of Kant’s successors were not convinced. Rather, it was by focusing even more intensely upon the nature of truth that several of them hoped to avoid the “subjectivizing” pitfalls of Kantian idealism. Bolzano, for example, turned his attention to the development of a revolutionary account of the bearers of truth (section 4.3); Brentano sought to shift the center of the discussion of truth toward a renewed and influential emphasis on our real experiences of correctness as definitive of truth (section 4.4); and Husserl hoped to supplement both Bolzano and Brentano’s analyses with a more direct and sustained analysis of the nature of truth-makers themselves (section 4.5).

In the process, these nineteenth-century theorists cast a critical spotlight upon each of the three dimensions noted above. The cumulative result was the cultivation of a dynamic philosophical context in which many of the key issues still at the heart of contemporary debates about truth were first identified as pivotal. Indeed, it brought about the very context in which the origins of two of the most influential movements in twentieth-century philosophy—analytic philosophy and phenomenology—find their roots. For this reason, even if thinking through the problems and prospects that emerge in this development does not, of itself, provide a complete resolution to these debates, it promises to provide us with both deeper clarity concerning these issues and a richer sense of the historical motivations for certain now-familiar theoretical twists and turns.

² Compare question 1, article 1 of Aquinas’s *Quaestiones disputatae de veritate* from the 1250s, as cited in Künne (2003: 102f).

4.2 KANT AND THE TRUTH IN APPEARANCES

Aquinas's definition was endorsed repeatedly throughout the early modern period, up till the time of Kant. We find it for example in Descartes, Spinoza, and Leibniz, as well as in two of Kant's more immediate predecessors, Christian Wolff and Georg Meier.³ We also find it explicitly endorsed by Kant himself in the *Critique of Pure Reason*: "the nominal definition of truth, namely that it is the agreement [*Übereinstimmung*] of cognition [*Erkenntnis*] with its object [*Gegenstand*], is here granted and presupposed" (B82). Kant affirms this definition at many points throughout the *Critique* and in several other contemporary writings as well.⁴ This suggests that Kant does not take himself to be putting forward a deeply revisionary or heterodox theory of truth, despite his clear sense that he is up to something revolutionary within theoretical philosophy, and despite his other radical departures from the Leibniz-Wolffian tradition. Even so, the threat to the familiar conception of truth becomes readily apparent once we look more carefully at Kant's own understanding of the three key terms in the traditional definition, especially his understanding of "object."

Concerning "cognition": though Kant above makes it sound as if truth can pertain to any species of cognition, his considered view is that truth is restricted to *judgments*: "truth as well as error . . . is to be found *only* in judgments" (B350; emphasis added). Judgments, like all cognitions, are a species of "representation [*Vorstellung*]," which means that a judgment is a mental act that is related to an object. Judgments are distinguished by being acts of our capacity for understanding—i.e. our "intellect"—acts, in particular, that involve the kind of synthesis or "combination [*Verbindung*]" of further cognitions that is expressed by the copula "is" (cf. B141).⁵

³ Cf. Descartes's October 16, 1639 letter to Mersenne (Descartes 1898: 597); (Spinoza 1677: part I, axiom 6); (Leibniz 1978 [1705]: Book IV, ch. 5); (Wolff 1740: section 505); and the textbook Kant used in his logic lectures (Meier 1752: section 99).

⁴ See B236, B196–7, B296, B848; see as well *Prolegomena* section 5 (4:279), JL section VII (9:50f), and the following student transcripts of Kant's lectures on logic: *Logik Dohna Wundlacken* 24:709; *Logik Busolt* 24:627; *Wiener Logik* 24:823. For a dissenting interpretation of Kant's "acceptance" of the tradition definition in this passage, see Prauss (1969).

In this chapter, I will cite major works parenthetically, according to the abbreviations provided in the References. For Kant, I will cite the first *Critique* according to the B-edition pagination, and will cite the *Prolegomena* ("Prol.") and Jäsche's *Logic* ("JL") according to the *Akademie Ausgabe* volume and page numbers (cf. Kant 1902–). For Brentano and Husserl, I will cite according to pagination from the English translation (where available) followed by that of the German edition. In all cases, the translations are my own, though I have consulted (and especially in the case of Kant's works, usually followed) the standard English translations, where possible (see References).

⁵ It is worth noting that, unlike Frege, "judging" for Kant is not equivalent to "holding-for-true [*Fürwahrhalten*]" (cf. JL section IX, 9:65f). Kant takes the act of judging to form a representation (a judgment) that is true or false (in the case of theoretical judgments), and which can then also *be held to be* true or false (in a separate act), but need not be.

Turning to “agreement,” we can note, first, that the particular representational “relation” that a given cognition bears to its object is what Kant calls the “content [*Inhalt*]” of the cognition (B79; cf. B83). Kant takes the agreement (and its opposite, the “contradiction”) between a cognition and its object to be a further specification of this relation, such that a particular cognition’s agreement with its object—i.e. its *truth*—is something that “pertains precisely to content” (cf. B83–4).⁶ In this respect, truth, for Kant, is primarily a semantical notion, as it concerns the representational relation between acts and objects, in virtue of their content. This contrasts with the more “metaphysical” view of truth, also anticipated in Aquinas, and put forward by some of the Leibnizians, according to which truth is a property of objects themselves.⁷ Kant explicitly rejects any conception that would make truth a “transcendental predicate of *things*,” rather than a property of our “*cognition of things*” (B113–14).

So far, so traditional. Things become considerably more complex, however, once we take a closer look at Kant’s mature conception of the third notion in the above definition of truth: the “object [*Gegenstand*]” to which our cognition “agrees” when it is true. For what Kant says about the objects of our cognitions threatens to eliminate their representation-independence altogether. Yet without this link to genuine objectivity, Kant’s views might seem to become deeply revisionary indeed.

Kant’s “Copernican” revolution comes about with his questioning of a commonly held assumption about the relation between our cognition and the objects of our cognition, with the hope that rejecting this assumption will allow us to finally make decisive progress in metaphysics. As he writes in the Preface to the *Critique of Pure Reason*,

Up to now it has been assumed that all our cognition must *conform to the objects* [*sich richten nach den Gegenständen*]; but all attempts to find out something about them a priori through concepts that would extend our cognition have, on this presupposition, come to nothing. Hence let us once try whether we do not get farther with the problems of metaphysics by assuming that the objects must *conform to our cognition*, which would agree better with the requested possibility of an a priori cognition of them, which is to establish something about objects before they are given to us. (Bxvi; emphasis added)

The first *Critique* is Kant’s attempt to explore and vindicate this inversion of the common assumption, with Kant’s conclusion being that the only metaphysical, a priori cognition that is possible is, in fact, of objects that “conform” to our cognition.

Along the way, however, Kant’s views about metaphysical cognition turn out to be intertwined with his account of cognition in general, whether metaphysical (a priori) or otherwise. The core of Kant’s argument about metaphysical cognition lies in his belief

⁶ As he puts it later in the *Critique*: truth “is to be found . . . *only* in the *relation* of objects to our understanding” (B350; emphasis added).

⁷ This view can be found in Baumgarten (1757: section 89), the textbook Kant used in his metaphysics lectures. For references to Aquinas’s discussion of “truth in things [*in rebus*],” see Künne (2003: 104).

that the only thing that our minds can be thought to have a priori access to, the only thing “given and present” to the mind a priori, is the mind itself, its own structure, and its two basic capacities for representing—“the constitution of our capacity for intuition” (our sensibility) and the “rules” for our understanding as a capacity for thinking and judging—since all of these are things that lie “in myself before any object is given to me” (Bxvii; cf. Prol. section 9, 4:282). Yet precisely because our a priori cognition is limited in this way, the only information that such cognition can convey about *objects* is that any object represented by these capacities will be represented in the forms that representations from these capacities must take, due to the nature of the capacities themselves. In this way, our a priori cognition of objects is restricted to knowledge about whatever *representations* of objects can “conform” to our capacities for cognition, rather than extending to cognition of these objects as they are “in themselves.”

The consequences for our cognition in general emerge once Kant begins to spell out what we know a priori about our mind. In particular, Kant thinks we know that an object can only be “given” to us in intuition if our mind’s sensibility is “affected” by it, such that we “sense” it in some way or other, with the effect of this affection being that the object is able to “appear” to us (B33). But we also know a priori that any representation (intuition) from our sensibility must “conform” to the way that this capacity represents things. For this reason, the appearance of an object in an intuition can only take place in a representation that will inevitably bear the marks of having been produced by our mental capacities. Even so, the universal and necessary presence of these marks in the appearance—what Kant calls “the forms of appearance,” and what he identifies with the spatial relations and temporal orderings that organize the contents given in our sensations—is something that we know is present due to the nature of our minds rather than the nature of the affecting objects (B34).

Now, Kant thinks that, a posteriori, it is ultimately the *appearances* of objects that are “the *only* objects that can be given to us immediately” (A108–9; emphasis added). These “objects,” however, are not really genuine self-standing things in their own right, since they are not a way something could be “in itself,” but are instead “representations” of things (A109). But then, because an appearance exists only as a “way of perceiving” or “representing” something else, Kant claims that it *only* exists “in us,” or at the very least, only in the relation between the object and our minds (B59; cf. Prol. section 52c, 4:341).

With this further claim, Kant can seem to straightforwardly threaten the mind-independence of the possible objects of our cognition *in general*, for the following reasons. If the only object that can be immediately present to us a priori is the mind itself (and its capacities), and the only things that can be immediately present a posteriori are appearances, and if the only things we can have knowledge of are things we can have immediately present before our minds, then our cognition, in general, will be “restricted” or “limited” to the mind and appearances. But since appearances are objects whose forms are supplied by our minds, then our cognition is, in general, restricted to objects that, in a very straightforward sense, *must* “conform” to our capacities for cognition, since they are either these capacities themselves or a “product” of them (Prol. section 20, 4:300).

With this we see why Kant's revolutionary account of our cognition and its objects can seem to have no grounds for retaining anything like the traditional notion of *truth*, and more specifically, the distinction between truth and falsity—Kant's own assent to the traditional definition notwithstanding. On the traditional picture of conformity or agreement of our representations with their objects, the former entities are beholden to the latter: the way objects actually are makes certain representations of them true, and the fact that we can represent objects otherwise than how they actually are makes falsity possible. On Kant's new picture, however, the objects of possible cognition are now beholden to our mental acts, as something of the mind's own making, as merely "ideal" rather than "real" (cf. B66). Yet with this inversion, the possibility of gaps between a putative object of knowledge *being* a certain way and its being *represented* as being that way can seem to have disappeared. The fact that our representations "produce" their immediate objects means that these objects simply do not have the sort of existence that outstrips their being represented; as Kant says repeatedly, they *are* nothing outside of their being represented. But then a conformity or agreement between cognition and object would seem to be guaranteed, universally and necessarily, and *all* of our intuitions (representations of appearances) would become trivially true.

At this point, however, Kant can insist that his position is being mischaracterized in two important respects. The first has to do with the nature of appearances. It is not *every* feature of appearances that is said to "lie in the mind" a priori, but rather only their *form*; this is part of why Kant thinks a better name for his idealism would be "formal idealism" (Prol. 4:337). Kant does not think that the "matter" of any appearance is itself already present as an a priori contribution due to nature of our mind's capacities. Rather, this matter is present only because of the contribution that the object makes to its appearance: it is present as "an effect of an object on our capacity for representation," and consists in what Kant calls the content of a "sensation [*Empfindung*]" (B34). This dependence of the matter on objects lends some independence to the appearances themselves and gives Kant a hook on which to hang his "empirical *realism*" about appearances (cf. A368f).

What is more, Kant himself concedes that the "agreement" of the products of a capacity for representation with the basic constitution of the capacity itself could achieve only "the form of truth," rather than full-blooded truth, since the complete cognition (form plus matter) could, at least in principle, still go on to "contradict" its object (B84). Hence Kant himself would admit that the mere "agreement" of appearances with the forms of cognition is at best a merely "*formal truth*," since this, in effect, "consists in the agreement of a cognition *with itself*" (JL section VII, 9:51; emphasis added). This contrasts with genuinely "*objective truth*," which requires the "matter" of the appearance, too, to "agree" with its object (*ibid.*).

This first appeal, however, can seem to help only so much. For one thing, Kant takes contents of sensations themselves to also be contained "in" the appearance, and hence also immediately present to the mind in a way that the *object* affecting us is not. This

leaves opaque the exact relation between these contents and whatever features of the object of the appearance they are representing.

Equally problematic is Kant's thesis that "*all* the properties that make up the intuition of [an object] belong *merely* to its appearance" (Prol. 4:289; emphasis added). This suggests that even the sensory qualities that fill in the forms supplied by the mind—even this "matter"—is ultimately something that belongs merely to the appearance of things. But if these material features of appearances are likewise things that do not exist outside of being represented, then they don't actually seem to possess the requisite representation-independence to function as facts that can constrain our representations to make them true.

Here, however, we must recall that Kant does not think that *appearances* are themselves actually the genuine bearers of truth and falsity. As we saw above, this title belongs instead to our *judgments*: "truth and illusion are *not* in the object, insofar as it is intuited"—i.e. in its appearance—"but in the judgment about it, insofar as it is thought" (B350). It is, therefore, only in judgments concerning appearances that any questions of agreement in the sense of truth arise.

What is more, Kant thinks we are able to make *true* judgments about the status of the appearances, judgments whose contents "agree" or conform exactly to the way their objects actually are:

[I]f I take all the [sensible] representations together with their form—namely, space and time—for nothing but appearances, and these last two for a mere form of sensibility that is by no means to be found outside it in the objects . . . then in the fact that I take them for mere appearances is contained not the least illusion or temptation for error . . . (Prol. 4:291)

We can make true judgments about appearances whenever we say exactly what Kant himself says about them and therefore we "take [*halten*]" them for what they really are—namely, appearances. Here the object of our judgment (appearance) is just as we are representing it to be in our judgment. Furthermore, at this point, Kant thinks that error is clearly possible; indeed, many metaphysicians prior to Kant wrongly "take" what belongs to the appearance of an object to belong instead to the object itself, and in this way "make mere representations into things" (Prol. 4:293).

Actually, Kant thinks we do even more than simply "taking" or "holding" appearances "for" what they really are. This is something he makes clear in the following important footnote in the *Transcendental Aesthetic*:

The predicates of appearance can be attributed to the object in itself in *relation* to our sense, e.g., the red color or fragrance to the rose . . . What is not to be encountered in the object in itself at all, but is always to be encountered in its *relation* to the subject and is inseparable from the representation of the object, is appearance, and thus the predicates of space and of time are rightly attributed to the objects of the senses as such, and there is no illusion in this. On the contrary, if I attribute the redness to the

rose in itself . . . or extension to all outer objects in themselves, *without* looking to a determinate *relation* of these objects to the subject and limiting my judgment to this, then illusion first arises. (B69–70n; emphasis added)

Above and beyond simply (and correctly) ascribing the property *being an appearance* to the immediate object of our intuition, we can also correctly ascribe the “predicates” that are contained “in” the appearance itself—i.e. all of the “properties that make up an intuition” (to recall a passage cited above)—not to the object as it is in itself, but instead to the object *in relation to* our capacity for sensing it. What is represented in these more sophisticated judgments will also be able to “agree” with their “object” because their object is now the complex: object-in-relation-to-my-sensibility; and we can see straightaway (thinks Kant) that *this* “object” is just as it *appears* to be.

With this, Kant would seem to have escaped the worst of the difficulties posed above. Since the object of the judgment is not dependent for its existence on the judging itself, it enjoys a form of relative representation-independence (i.e. relative to judging, even if not relative to intuiting). This, moreover, also opens up space for our judgments to misrepresent these objects, since we are “entirely free” to “take” them to be something other than they actually are, as when we take them to be, or contain properties of, things in themselves (cf. Prol. 4:290).

Yet however much is clearly gained for Kant’s position by returning our focus to *judgment* rather than appearance, this shift brings with it a further difficulty—a problem, moreover, that is perhaps the deepest yet encountered. Recall that judgments arise through acts of combination, combination that is expressed by the copula “is.” Now, Kant takes this combination to be the result of an “act [*Actus*]” of our capacity for understanding, something “executed by the subject itself” out of its “spontaneity” or “self-activity [*Selbsttätigkeit*]” (B130). Importantly, for Kant, this means that such combination is a feature of our representations that “can *never* come to us through the senses” (B129; emphasis added) and “is *not* given through objects”; rather, Kant thinks we “cannot represent something as combined in the object without having previously combined it ourselves” via our understanding (B130; emphasis added).

From this, however, it would seem to follow that the combinatory form of what is represented in judgment is not a feature that tracks anything *either* in the objects as they are “in themselves” *or*—and this is the crucial point—“in” objects as they *appear*. For, just as the fact that spatial relations and temporal order are put into intuitions by the exercise of our sensible capacity implies that they are not representative of something present in the object “in itself,” so too should the fact that predicative combination is put into our judgment-representations only by acts of our conceptual capacity (understanding) entail that *it* is not representative of something present in the object of the judgment (the thing in relation to my sensibility through intuition (appearance)). But then if we know that the very form that judgment must take is not something that will “agree” with *any* object—either objects as they are in themselves, or objects as they appear (appearances)—all routes would seem to be blocked for making the case that what is represented in the judgment, *as a whole*, could somehow stand in agreement

with—and hence, be “true” of—either of these objects nevertheless. And, a fortiori, it is unclear what it would mean for us to claim to “see” this agreement or have evidence for taking it to obtain (and so for holding a judgment to be true).⁸

Though the force of this predicament concerning the lack of objective correlates for the very form of judgment was recognized by some of Kant’s immediate successors (perhaps most notably, Hölderlin and Hegel),⁹ a more sustained attempt at its resolution was not undertaken until several decades later, when Brentano and his followers wrestled with the notion of distinctive objective correlates for whole judgments—what are now most commonly called “states of affairs” or “facts”—and how they might be “given” to us. In the meantime, however, a challenge to Kant’s account of truth arose from an altogether different angle.

4.3 BOLZANO ON THE OBJECTIVITY OF THE BEARERS OF TRUTH

Bolzano belonged to a counter-stream in post-Kantian German-language philosophy, one that paralleled chronologically the Idealist and Romantic developments, but which took such developments—along with Kant’s original turn toward transcendental idealism and his “subjectivist” emphasis on the mental capacities—to have set philosophy off on the entirely wrong track.

In his *Wissenschaftslehre* (“Theory” or “Doctrine” of Science), his 1837 masterwork, Bolzano places one particular failing at the center of his critique of the Critical philosophy: the failure to think carefully enough about the nature of truth itself, and in particular, a failure to identify the genuine *bearer* of truth. Bolzano is in broad agreement with Kant in holding that the bearer of truth is representational in nature. Hence, like Kant, Bolzano too rejects the application of the property of being true directly to objects, simply in virtue of their *being* or existing; this sense of “truth” Bolzano also calls

⁸ Kant concedes that it is not in virtue of seeing that any such “agreement” obtains between an appearance and its object that we take a cognition to be true. For despite taking the notion of agreement to provide the content of the concept of truth, Kant rejects the idea that this notion can function as an independent *criterion* for use in telling which judgments are true. Knowledge of such agreement would just *be* the knowledge of the truth of the judgment, rather than an apprehension of a mark that could *then* be used as a sure sign for taking something to be true (cf. JL section VII, 9:50). In fact, Kant accepts that the only criterion that we have for holding a judgment to be true, prior to knowing it to be true (to “agree”), is the internal “coherence [*Zusammenstimmung*]” of our claims about what appears with our other claims about appearances (B179; cf. Prol. 4:290). At one point Kant even seems to accept that such coherence is a “*sufficient mark*” for truth relative to appearances (B679; emphasis added). For interpretations which claim that Kant already means to embrace a more coherentist *definition* (rather than criterion) of truth, see Windelband (1884); Kemp Smith (1918); Walker (1989). For a discussion of the difficulties facing this sort of reading, see Van Cleve (1999).

⁹ Cf. Hegel (1812: Introduction); for some discussion of Hölderlin, see Henrich (1997).

“transcendent” or “metaphysical,” as opposed to its genuine “logical” sense, which is linked to judgments (WL section 27, I.118f). Yet Bolzano also thinks that describing the bearer of truth as a kind of representation tends to cover over an important ambiguity in the term “representation” itself—namely, the ambiguity between its picking out the *act* of representing and its picking out the *content* contained in such acts. In fact, Bolzano takes the failure to “distinguish sharply enough” between the act and the content of our representations to be “the source of most of the current errors in logic”—Kant’s included (WL section 12, I.47).

Even if (as we saw above) Kant had noted that acts of cognition possess a content, construed as a representational relation to an object, in addition to this object itself, Bolzano thinks Kant did not do nearly enough to clarify this distinction, nor did he undertake any sustained investigation of the nature of these contents directly. What such investigations show is that the contents of our representings possess an identity that is *independent* of the reality of any one of these acts. The very same content can be taken up in multiple acts at multiple times (without itself being “multiplied”) or might never be thought of or grasped by anyone at all—save, perhaps, by God (WL section 48, I.217–18). But then such contents can in no way be products, effects, or creations of these acts either (cf. CE 115; 32 and CE 142; 63). Consequently, Bolzano thinks we should regiment our terminology and speak of “subjective representation” when we mean to pick out real mental acts that exist in some subject, and use “objective representation” or “representation in itself [*an sich*]” when we mean to pick out the act-independent self-identical content (or “matter [*Stoff*]”), which is “not something existing” and “is not to be found in the realm of the actual” (WL section 48, I.217).

Bolzano takes this threefold distinction between act, content (“matter”), and object to apply at the level of judgments as well. Here he introduces the terms “objective proposition [Satz]” or “proposition in itself”—for short, simply “propositions”—to pick out the content of judgments (WL section 19, I.76–8). Since propositions are themselves composed of objective representations (WL section 48, I.216), they, too, cannot be ascribed “a being [*Dasein*] (existence or actuality [*Wirklichkeit*])” at all (WL section 19, I.78).¹⁰ Rather, just like their constituents, propositions have an identity that is distinct in kind from the acts in which they are grasped, and they are not brought into being by any act of mind either. In these respects, Bolzanian propositions are closer kin to Fregean “thoughts [*Gedanken*]” than they are to Russell’s propositions.¹¹

With this in mind, Bolzano argues that the genuine bearer of the “property [*Beschaffenheit*]” of *truth* must be the objective content of such an act—i.e. the proposition—rather than the (subjective, real) act of judgment (WL section 24, I.108). Though this partially echoes Kant, Bolzano distances himself from Kant by insisting

¹⁰ Even so, Bolzano thinks that (like objective representations) propositions are objects (WL section 25, I.115), and that the proposition expressed by “there are [*es gibt*] propositions and truths” is true (WL section 30, I.144).

¹¹ Though it is worth noting that Bolzano uses “Gedanke” itself in an un-Fregean way, to refer to mental acts of thinking (cf. WL section 19, I.78).

that the content of a judgment itself is not something that is “combined” or put together by any “act” (spontaneous or otherwise); rather, any combination that is present in a proposition is present in this content “in itself,” as it were.

Still, Bolzano does accept that propositional content contains a combination of sorts. Bolzano takes every proposition to consist in three parts: a subject-representation which represents an object, a predicate-representation which represents a “property [*Beschaffenheit*],” and the concept of “having [*Haben*]” (which Bolzano prefers over “being”) that functions as the copula or the “connecting part [*Bindeteil*]” that links the representation of the property to the representation of the object (WL sections 126–7, II.9). Bolzano typically represents this structure as “A has b.” It is this link that makes the property be “asserted of this object” in the proposition (WL section 28, I.122), though this asserting, of course, is something that propositions “do” in themselves, so to speak, independently of anyone actually having them in mind (WL section 24, I.114).

What then does Bolzano think distinguishes true propositions from false ones? He claims that propositions are true when “they assert [*aussagen*] something as it actually [*wirklich*] is” (WL section 24, I.114). More specifically, propositions are true when they assert a property of an object that, in fact, belongs to the object: “in a true proposition, that which is *asserted* [*ausgesagt*] of the object must *actually pertain* to it [*wirklich ihn zukommen*]” (WL section 28, I.122; emphasis added). In false propositions, by contrast, this coordination between asserting and pertaining is lacking: there is the “mere asserting [*blosse Aussagen*]” of such a “belonging or pertaining” without such a connection “actually” obtaining (section 28, I.123).

This appeal to a coordination between content and object might seem to bring Bolzano’s account of truth in line with the traditional analysis in terms of *adaequatio*. Yet when Bolzano takes up the traditional definition itself, he claims not to see what is gained by using the terms “correspondence” or “agreement” to refer to such a coordination (cf CE, 167; 90; WL §29, I.128; and WL §42, I.180). Even so, he applauds Aristotle for claiming (in *Prior Analytics* I.1) that “the following two manners of speaking are *identical* [*identisch*]”: “this pertains to that” and “this can in truth be asserted of that” (WL section 28, I.124; emphasis added). And however it is labeled, Bolzano does think that to say that such a coordination obtains—and, hence, to assert of a proposition that it is true—is to say something *substantive*: “*p* is true” predicates a genuine property of an object—namely, the ideal object that is the proposition in question.¹² Bolzano’s

¹² In the note to WL section 32 e.g. Bolzano distinguishes between the proposition “A is B,” and “the proposition, that A is B, is true,” on the grounds that the latter “is a different one according to its component parts, and thus a second truth distinct from the first” (I.147). At the very least, the two propositions have different subject-terms: for any proposition “A,” we find that the proposition expressed by the words “A is true” is one distinct from the proposition “A” itself, since the former obviously has a completely different subject from the latter. Its subject is, namely, the complete proposition “A” itself (Bolzano 1851: section 13, 13). This is so, despite the fact that, as Bolzano acknowledges, “if the proposition ‘A is B’ is true, then so too the assertion ‘the proposition, that A is B, is true’ is a true proposition” (WL section 32, I.147)—i.e. despite the fact that the semantical predication itself *follows* from the truth of the original proposition.

view, then, stands in direct opposition to redundancy theorists (and also to Frege, on some readings), even if he agrees that what *can* be said about the nature of this property is very limited.

Yet in light of the previously encountered difficulties concerning what in the object is to correspond to the copula in a true judgment, Bolzano's reticence about the right way to characterize the relationship between true propositions and the objects they represent might begin to look more problematic. Worries deepen once we draw out Bolzano's account of propositions. Being themselves representational in nature, propositions (like all objective representations) must be distinguished from the "objects" to which they are "related" representationally (cf. WL section 49, I.218f).¹³ What Bolzano explicitly identifies as the "object [*Gegenstand*]" of a proposition as a whole, however, is simply the object of the subject-representation of the proposition; at certain points, Bolzano even calls the subject-representation simply the "object-representation" (cf. WL section 126, II.8).¹⁴ Yet Bolzano clearly accepts that there is *more* that is being represented "in" a proposition than simply the object. Recall that, for Bolzano, propositions possess the following kind of structure: A has b. Hence, there is not only the subject-representation, but also the predicate-representation that represents a property, and, in addition, the copula or "linking part" or "linking member" that "indicates [*anzeigt*] that the object *has* the given property" (WL section 126, II.8–9).

Now, if what the whole proposition were to be coordinated with was simply the object of its subject-representation, then this would seem to make the rest of the propositional representation superfluous for its truth, since the relevant coordination would have already been taken care of simply by the subject-representation itself. What the proposition as a whole seems to aim to coordinate with, then, is instead something like the fact that A has b, or the having-b of A. That Bolzano is angling for such a view would seem to be further encouraged by the willingness we saw above to identify b's (actually) "pertaining to" A as that which is coordinated with b's being predicated or "asserted of" A in the proposition, when that proposition is true.

The problem with crediting Bolzano this line of analysis, however, is that, like Kant, Bolzano ultimately explicitly denies that there is anything at the level of objects and properties that corresponds to what is represented by "has." In WL section 78, Bolzano claims explicitly that "has" belongs to the class of representations "that have no *object* at all," along with 'nothing', 'and', and 'round square' (I.360).¹⁵ And it follows from his analysis in WL section 64 that "has" does not pick out a property either. There he emphasizes that, in composite representations, in addition to the representation of the object and the

¹³ Bolzano thinks that, in every case, a content must be distinguished from the *object* that it represents, both because distinct contents can represent the same object, and because—as is especially evident in cases when the object is something really existent—the content and its object bear obviously distinct properties (WL section 49, I.219).

¹⁴ In a similar fashion, Bolzano identifies the "extension [*Umfang*]" of a proposition as a whole with the extension of the subject-representation (cf. WL section 130, II.25).

¹⁵ In other words, it is an "objectless [*gegenstandlos*]" representation (cf. WL section 67, I.304).

“mere representations of its properties,” there will “still be need of some *other* representations as well which will serve to combine [*verbinden*] them” (emphasis added):

In order to represent, in particular, that the object has the properties *b*, *b'*, *b''* . . . in itself, one must form [*bilden*] the representation: “something which has (the properties) *b*, *b'*, *b''* . . .” In this representation, however, there are many other representations besides the representations of the properties *b*, *b'*, *b''* . . . —namely, the representation of *something*, the representation of the relative pronoun *which*, and the representation of *having*. (WL section 64, I.270–1)

Here Bolzano claims explicitly that the concept of having, i.e. the copula, represents something “besides” any of the properties of the object. But then, while the properties *b*, *b'*, *b''* and so on are all properties that *A* has (properties that pertain to *A*), it follows that neither the *having* of *b*, the *having* of *b'*, and so on, nor (so it would seem) *b*’s *pertaining* to *A*, etc. are themselves further properties that *A* has.

For Bolzano, then, the only significance of the copula would seem to be its sense-conferring role of “linking” the other representations into something that has the form of a proposition. Rather than itself representing any object or property, the purpose of the representation expressed by “has” appears to be fulfilled entirely internally to the economy of representations, by serving to “combine [*verbinden*]” the other (objective) representations. Yet if its presence in a proposition is *not* demanded of it by its *object*, then why is the combination present in the proposition, in the sphere of representations, in the first place?

The traditional answer up through Kant was that such combination is present in the bearers of truth, not because of the objects, but as a result of mental activity.¹⁶ Now, as we saw above, this also implied that the copula does not have an objective correlate (and is hence “*gegenstandslos*”). Indeed, it was precisely this feature of the view that posed an obstacle to analyzing truth as an agreement between contents of judgments and the things represented. To be sure, this act-theoretic account of the significance of the copula is unavailable for Bolzano, since he denies that combination is present in propositions as the result of mental activity. But then, by rejecting both the mind-dependence of propositions *an sich* (or their forms), as well as any account on which their form is made to order, as it were, as a result of something “in” their objects, Bolzano appears to leave us with no explanation whatsoever of the distinctive unity and structure of propositions.

Furthermore, by accepting the presence of representational *content* in propositions (and hence, in truths themselves) that in no way tracks the way *objects* themselves are *an sich*, it is hard to see how Bolzano’s account will be able avoid re-introducing Kant’s “subjectivizing” distinction between the way that objects “appear” to us—even if now in

¹⁶ In Locke (1979 [1689]: Book III, ch. 7), Locke for example gives a similar account of the significance of “is” and other “particles”: such words “signify the connection that the mind gives to ideas” (section 1), and thereby serve as “marks of some action, some intimation of the mind” (section 4). In Arnauld and Nicole (1992 [1683]: Part II, ch. 2), the Port-Royal logicians similarly describe the “principle usage” of words like “is” as that of “signifying movements of our soul.”

“objective” propositions “in themselves”—and the way *they* (the objects) are “in themselves.” In effect, Bolzano seems to trade Kant’s sensible idealism for a “logical” idealism.

This is especially troubling, given Bolzano’s additional claim that all of our “knowledge [*Erkenntnis*]” of objects takes the form of grasping true *propositions* about them in judgment (cf WL section 36, I.163). Insofar as the very contents of our ostensible knowledge are not transparent guides to how things stand in the realm of objects, Bolzano seems unable to escape Kantianesque conclusions concerning the restriction of our knowledge to the (now) propositional “appearance” of an object. So long as this gap remains, Bolzano’s account of propositions, and therefore truths, threatens to leave us one step short of the facts, and therefore to eliminate the intelligibility of the first aspect of the traditional notion of truth identified above—namely, truth’s beholdenness to the *res*.

At the same time, however, Bolzano also threatens to eliminate the third aspect from the traditional understanding as well—namely, its essential link to an *intellectus*, to the mental activity of representers (subjects) like ourselves. We have already seen that Bolzano takes propositions to possess an identity and a structure that is what it is independent of any relation to any actual existent mental activity. In fact, Bolzano thinks that the nature of propositions is also fixed independently from any relation that they might bear to any *possible* mental activity as well. This can be seen from his claim that even if it is true that God can (and does) think every proposition and can (and does) know every truth—and hence, even if it is true that the *properties* of being thinkable and knowable belong to every proposition and truth as *objects*, respectively—that even so, the *concepts* of a proposition and of a truth do not “include” or “contain” the *concepts* of being thinkable and being knowable (cf. WL section 23, I.92, and section 25, I.113).

Hence, despite the fact that Bolzano himself first introduces these concepts by pointing to their function as the content of subjective representations, he ultimately takes the concept of a proposition and a truth to be both concepts that are intelligible independently of the concepts of mental acts. But if *these* concepts are not included in the concept of being a proposition, it is hard to know what such a concept *does* contain. Indeed, insofar as propositions are essentially representational in nature, one might wonder what else could belong essentially to the concept, if not the notion of being something that can represent an object (as being a certain way) *to a mind or subject*?

4.4 BRENTANO AND THE EXPERIENCE OF TRUTH

Brentano saw quite clearly, from early on in his career, that the problems opened up by Kantian idealism would not be blocked completely and decisively until a better account of the place of combination in the contents of judgment was provided. Even so, as we will see below, throughout his life, Brentano remained highly self-critical of each of his own previous attempts to resolve these questions. This led to a very rich progression of views

through roughly three stages of Brentano's writings: an early period, in which he explores what might be called a more "metaphysical" conception of truth, which makes the bearers of truth the objects themselves; a middle period, in which he takes a turn toward the more traditional, "logical" conception of truth, in which judgments are restored as the truth-bearers, though with an innovative conception of the objective correlates of such judgments; and a final period which sees Brentano focus instead on certain primitive experiences that we have as the key to understanding truth.

There is, however, one very important commitment that Brentano embraced early on and never relinquished, a commitment that—in light of the preceding sections—might seem to be a natural option for someone attempting to avoid the pitfalls of idealism. Nevertheless, it was a quite radical one, from the traditional perspective of thinking about judgment, especially after Kant. This is Brentano's unfailing rejection of the idea that the copula in judgment consists in, or represents, *any* sort of *combination* at all.

Brentano presents his case against this traditional doctrine toward the end of his early and most well-known work, the 1874 *Psychology from an Empirical Standpoint*. He points out, first, that some cases of mere representing without judging have complex objects (e.g. merely thinking of "a green tree"); this Brentano takes to imply that combination among representations is not sufficient for judgment (PES 205; II.45). He then argues, secondly, and more importantly, that, even in judgment, combination is not necessary: "affirmation and denial are not exclusively directed at composites or relations," since a single object, or even "a single feature of an object," is such that it "can be affirmed or rejected" (PES 208; II.48–9). Brentano takes a paradigmatic case of this to be found in existential judgments of the form "A is" and "A is not." Concerning such judgments, he writes: "It is not the combination of an attribute 'existence' with 'A', but 'A' itself is the object that we affirm . . . The affirmation of A is the true and full sense of the proposition, therefore nothing other than A is the object of the judgment"—with the complementary point being made about the negative form (PES 208–9; II.49–50).

What is it, then, that Brentano thinks distinguishes these existential judgments from mere representings? Brentano focuses his analysis, unsurprisingly, on the presence of the "is." He agrees with Kant and Bolzano that the existential "is," like the "is" in general, does not have an objective correlate in the ordinary sense: "the 'being [*Sein*]' of the copula does not signify [*bedeute*] anything of itself, as a name does" (PES 212n; II.54n). Rather, like Kant, though now unlike Bolzano, Brentano holds that this "is" signifies an act of the mind: "it completes the expression of an act of judgment" (*ibid.*). Yet contrary to Kant's official view, Brentano thinks that it is obvious that the relevant act consists *not* in a synthesis or combination of contents with other contents, but rather in a simple *thesis* or act of positing or "affirming [*anerkennen*]," or, in the negative case, an act of a simple "negating [*leugnen*]" or "rejecting [*verwerfen*]" (cf. PES 210f; II.53f).¹⁷

¹⁷ Brentano recognizes that Kant's well-known doctrine that being is not a real predicate points in the direction of this sort of analysis of existential judgments, though he thinks that Kant failed to follow through on this insight, insofar as Kant still maintains that existential judgments are *synthetic* (cf. PES 211; II.53). For discussion of the relation between Brentano and Kant on this point, see Martin (2006: ch. 2).

Brentano's next claim, however, is the most revisionary. Far from simply offering one isolated counterexample to the traditional doctrine of judgment, Brentano insists that the structure of existential judgment provides all we need for "nothing less than a complete overthrow" of the traditional doctrine of judgment, and hence, "at the same time, a reconstruction of elementary logic" (PES 230; II.77). This is because Brentano thinks that *all* expressions for judgments can be "reduced" to "sentences" in existential form. Brentano takes this to be true even of sentences that appear to express judgments involving acts of combination or predication, such as those expressed in categorical "sentences" like "All A is B," and also of those involving even more complexity, such as those expressed in hypothetical "sentences" like "if A is B, then A is C." The former becomes "A non-B A is not" (cf. PES 214; II.56–7), while the latter becomes, first, "All AB are AC," and then (like the former) "A non-AC AB is not" (cf. PES 218; II.59). Granting Brentano's assumption—not implausible at the time, in light of the then-contemporary doctrines of pre-Fregean logic—that all expressions for judgments are, at root, in either categorical, hypothetical, or existential form, Brentano takes himself to have shown that "the reducibility [*Rückführbarkeit*] of *all* sentences which express a judgment to existential sentences is indubitable" (PES 218; II.60; emphasis added).¹⁸

With this Brentano has rejected Kant's idea that any special mental act is necessary for combination (complexity) to be represented. Rather, he thinks that combination *can* simply be "given" straightaway, present to the mind in an act of mere representing. Indeed, on Brentano's new picture, the very same object, whether simple or complex, is both the possible target of a mere representing and of the correlative act of judgment:

With regard to content, there is not the slightest difference between [judgment and representation]. The same object is had in consciousness whether a person affirms it, denies it, or is uncertain and merely questioning it; in the last case the person is merely representing it, in the first two cases it is being both represented and affirmed or denied. (PES 221; II.63)

Judging is to be distinguished from "mere representing," therefore, not by the introduction of a new more complex kind of objectivity or content, through acts of combination, but instead solely by the "particular kind of mental relation [*Beziehung*]" to the original object or content of the representation, whether that object or content is simple or composite (PES 240; II.89). It is one and the same "object of representation" that "*becomes* the object of an affirmative or negative judgment" when "our consciousness enters into a completely new kind of relation [*Art von Beziehung*] to the object" (PES 201; II.38; emphasis added).¹⁹

¹⁸ For more on Brentano's proposed revision of the traditional logic, see Simons (1987; 2004).

¹⁹ As Anton Marty, one of Brentano's students, noted as early as 1884, this brings Brentano's views on the relationship between judgment and mere representation quite close to Frege's account in section 2 of his 1879 *Begriffsschrift* of the relationship between the act of judgment and what is expressed by the horizontal content-stroke—though, as Marty also notes, Brentano retains a duality at the level of acts (affirmation and denial), whereas Frege's account moves this distinction into the content judged (cf. *Begriffsschrift* section 4), leaving only the single act of asserting (cf. Marty 1884: 185f).

Brentano intends for his account to be anti-idealist in another fashion as well, insofar as he insists that, in cases of adequate perception, when the object is given, it is genuinely the *object* that is given to the mind, and not merely some intermediary content of a representation of it.²⁰ This can be seen in Brentano's *identification* of the contents of representations with their objects, that is, his identification of the act's "relation [*Beziehung*] to a content [*Inhalt*]" with its "directedness [*Richtung*] toward an object" (PES 88; I.124). In this respect, Brentano appears closer to the direct realism of the early Russell and Moore (himself a close reader and admirer of Brentano) than he does to Bolzano or Frege.

Unsurprisingly, these revisions of the traditional conception of judgment bring with them a revisionary conception of *truth*. For though Brentano concedes that there is a link between judgment and truth, and though he, in effect, accepts that it is the contents of judgments that are the bearers of truth (cf. PES 141, I.200; 198, II.34; 223, II.67; 239–40, II.88–9), because he does not sharply distinguish between the content and the object of the judgment, this ultimately leads him to take *objects* themselves, as contents of judgments, to be true. Thus we find Brentano frequently identifying both what is accepted (or rejected) as true or false, as well as what is true or false, with the *object* itself, even going so far as to talk at times directly of "the truth of the object [*die Wahrheit des Gegenstandes*]" (PES 240; II.89–90; emphasis added).²¹ On this more "metaphysical" conception of truth, there is no room for a gap between how something objective is represented as being—e.g. via some combinatory element that turns out to be "merely" in the content—and how it is "in itself." Importantly, however, this gap has not been eliminated due to the mind's acts "producing" the object.

By the late 1880s, however, Brentano's conception of truth had undergone a further shift, one in which he moves away from this metaphysical conception, and at least a few steps back toward the more traditional, "logical" conception favored by Kant and Bolzano—though with important modifications. Brentano presents the new position in his 1889 lecture, "On the Concept of Truth." He appears to have been led to revise his position by rethinking his account of negative existential judgments: cases in which the denial of the existence of an object is true. Such a judgment is true when the object is non-existent. Yet what *makes* this judgment true cannot *be* the object itself, since it does not exist. There must, therefore, be a distinct entity or objectivity that performs this service. Brentano thus now takes the complex *the non-existence of the object* as a whole to be the correlate of such judgments (cf. section 41 et seq.; TE 19f, 22f).

Accordingly, Brentano now defines *truth* in the following manner: "a judgment is true if it asserts, of something that is, that it is, and also when it denies, of something that is not, that it is" (section 51, TE 21, 24). Note, first, that Brentano is now conceiving of judgment as the affirmation or denial, not simply of an object, but instead of its existence

²⁰ Brentano thinks that these cases are limited to our inner, "incidental" perception of mental acts; cf. PES 19f and 30; I.28 and 41.

²¹ Here again there are parallels with the early Russell and Moore, to their so-called "identity" theory of truth.

or non-existence—something he describes in fact-like terms (“*that* it is”). Note, second, that what it means to say that a judgment is *true* is cashed out in terms of the correlation between the affirmation of the existence of the object and the existence of the object itself, and between the denial of existence and non-existence.

Brentano’s embrace of this new correlation is stated more directly in a supplement to his 1889 work, *On the Origin of Our Moral Knowledge*:

The concepts of existence and non-existence are correlatives to the concepts of the truth of (simple) affirmative and negative judgments . . . If I say that an affirmative judgment is true or that its object is existent, *in both cases I say precisely the same thing* . . . The assertion of the truth of the judgment, that there is a learned man, is thus the correlate of the assertion of the existence of its object, “a learned man”. (TE, 39n, 45n)

To say “*p* is true,” where *p* is affirmative, is to say “the object represented in *p* exists,” and what *makes* an affirmative judgment true is the existence of the object, rather than simply the object itself. And, finally, what *is* true is the judgment expressed by “*p*,” not its object.

With the introduction of these correlates and the parallel interest in the question of truth-makers, Brentano is also now willing to endorse a revised version of the traditional definition of truth, as the “agreement [*Übereinstimmung*]” or “harmony” of a judgment with its object (section 52, TE 21, 25). Revision is necessary, though, because what is at the heart of this “harmony” between judgment and object is actually a more sophisticated “correlation,” consisting in “the agreement of a true judgment with its object or [rather] with the *existence* or *non-existence* of its object” (section 54, TE 22, 25).

Brentano’s introduction of special object-like correlates for judgments—items that are both distinct from objects or things or realities in the typical sense and yet what make judgments true—marks a decisive shift in the theorizing of truth in the nineteenth century. This feature of Brentano’s middle-period views was extraordinarily influential among his students, colleagues, and readers. In decades to follow, many philosophers in Brentano’s circle of influence introduced very similar (though not identical) notions, such as Alexius Meinong’s “objective,” and Carl Stumpf’s notion of a “state of affairs [*Sachverhalt*],” not to mention Moore and Russell’s notion of a “proposition.”²²

This new account of truth introduces a compositeness into the “objects” that are involved in even otherwise simple existential judgments, one that was missing from the point of view of the *Psychology*. What is more, it introduces a distinction between an object and its being or existence, something the earlier Brentano was loath to do (cf. PES 240; II.89). Even so, Brentano retains one key feature of his revisionary account of judgment, insisting that whatever compositeness is present in these new correlates (the existence or non-existence of the object) is still not the result of their having been “combined” or “synthesized” by the mind in judgment, but are simply presented directly

²² For discussion and comparison, see Smith (1989) and Benoist (2006).

before the mind in mere representation, and then affirmed or denied (cf. section 30 et seq.; TE 16f, 18f).

In any case, much to many of his students' chagrin, Brentano's own acceptance of these distinct objectual correlates for judgments was short-lived. By the early 1900s, he had become deeply dissatisfied with his new position and reverted back to views much closer to those of the *Psychology*. The two main vices that were associated with the introduction of these states of correlates were the resulting bloated ontology and the absence of any plausible account of how we could be acquainted with such entities.

Concerning the first, Brentano thinks that embracing such correlates entails the absurd conclusion that "there are not merely the real things [*die realen Dinge*], but also their being [*Sein*] and non-being and a legion, indeed an infinity, of impossibilities" (TE 82, 91):

Anyone who says that, in addition to things, there is the being of things, as well as the non-being of things is also committed to this: in addition to the individual dog, there exists not merely the being of that dog, but also the being of each of its parts, however small, as well as the being of the limits which belong to it as a body . . . And then the being of the being of the dog, in turn, would require analogous assumptions. An infinite and entirely unfruitful complication. Yet the adventures one encounters with the non-being of an individual dog—whether one denies or affirms the dog itself—would be even more bizarre. (TE 111, 126)²³

Concerning the second, Brentano despairs of answering the following question: "how does one arrive at *knowledge* of the *being* of A, as distinguished from the knowledge simply of A itself?" (TE 108, 122). He takes the two options to be immediate apprehension through perception or intuition and mediate apprehension through inference or abstraction. The first is ruled out of hand:

Is [being] something that is immediately perceived? Is the being of this being, etc., also perceived at the same time . . . ? No one would assert [this], for the consequences are much too paradoxical . . . Intuition [*Anschauung*] of the being of A, alongside that of A, can clearly not be spoken of. (TE 108, 122)

Brentano's arguments against acquiring knowledge of such *entia* via inference and abstraction are equally quick: it is entirely unclear what premises could yield these *entia* as inferential conclusions; and abstraction would have the unfortunate consequence

²³ This is an argument that Brentano returns to again and again in his later letters and writings: cf. his 1906 letter to Marty (TE 86, 95–6), his 1914 letter to Kraus (TE 98, 110–11), a 1914 fragment (TE 108, 122), and a 1917 manuscript (PES 338; PES II.236). It should also be noted that the parallels between this *reductio* and Bolzano's entirely serious argument for the existence of an infinity of truths (present in both the *Wissenschaftslehre* and the *Paradoxes*) were not at all lost on Brentano; see especially TC 32, 29–30. Indeed, Brentano was happy to lump Bolzano in with Marty and the others as someone who also "ascended too high and lost his way" with the "absurd" supposition of "a realm of *Gedankendinge*," as Brentano puts it in a 1905 letter to Husserl (TE 137, 157).

that the concept of the being of A would have to be a “more general concept” than that of A itself, and so somehow include A under itself as a species (TE 108, 122). Having taken himself to exclude all options, Brentano concludes that no story about our representational access to such objects is forthcoming: “it appears to be obvious, therefore, that there can be no talk of a *representation* of the being of A in the proper sense, but instead that it is always *only* the A that we are representing” (TE 109, 123; emphasis added)—with the same thing following for the non-being of A as well (cf. TE 112, 128).

By 1904 at the latest, then, Brentano holds, not just that “there is nothing other than something real,” but also that “nothing other than something real can be *thought*” (TE 67–8, 78–9; emphasis added).²⁴ Yet however compelling the arguments against such *entia* may be, and however appealing the new austere realist (“reistic”) ontology may itself be, Brentano is now faced with making sense of *truth*, and the relation of making-true—and in particular, in the case of negative existential judgments—while avoiding reference to the very things he previously took to be necessary for such a task. How does Brentano now propose to deal with truth and falsity, if not in terms of the harmonious correlation, or lack thereof, between the quality of judgment (affirmative or negative) and the being (existence) or non-being of its object?

Brentano’s final proposal is, once again, quite radical:

Truth pertains to the judgment of the person who judges correctly [*richtig*—i.e., to the judgment of the person who judges about a thing in the way in which anyone whose judgment was made with *evidence* would judge about the thing; hence it pertains to the judgment of one who asserts what the person who judges with evidence would also assert. (TE 122, 139)

Rather than an analysis of truth in terms of the agreement between judgment and *object* altogether, Brentano now moves toward an analysis in terms of the agreement between a judgment and *another judgment*—namely, agreement with what is *evidently correct* to judge. Instead of consisting in the agreement of a judgment with an object toward which it is intentionally directed, truth is now taken to consist in terms of an agreement with what a judge who sees the evident correctness of his judgment would experience or perceive.²⁵

Once this step is taken, Brentano now thinks that he can provide a new analysis of negative existential judgments as follows. He argues, first, that concepts like “the

²⁴ As he indicates in a later manuscript from 1914, what Brentano means by “real” here includes things like “a body, a mind, a topoid of more or fewer than three dimensions” (TC 16, 4), and more generally “every substance, every plurality of substances, every part of a substance, and also every accident” (TC 19, 11). Brentano also identifies “the concept of what is, in the proper sense” with “the concept of what is temporally present [*das Gegenwärtige*]” (TC 20, 12; cf. TC 24, 18). Only such “*realia*” are “what is in the proper sense [*im eigentlichen Sinne*]” 1914 (TC 16, 4).

²⁵ Brentano sees this later view as harking back to the explanation of truth in terms of clear and distinct perception (“*verum quod clare et distincte percipio est*”) that Descartes had elaborated in the Third Meditation (TE 124, 141).

existent” and “the non-existent” appear to be correctly applied to objects because the object in question is either “something correctly thought in a positive manner” (“something correctly affirmed”) or something correctly thought in a “negative” manner (TE 68–9, 79). Yet Brentano now notes that though properties like “thinking of X correctly” *look* like they are relational properties, they are really species of monadic properties of the form *thinking in a certain way*.²⁶ This opens up a path for taking kinds of representing or thinking that *look* like they involve reference to non-existent objects—whether of the mundane sort (e.g. a centaur), or of the more exotic, philosophical sort, like the existence or non-existence of a dog—and re-analyzing them as consisting instead in thinking or representing in certain ways. This, in effect, leads Brentano to a reconstrual of intentional content along the lines of what now goes under the name of adverbialism, which, in turn, is meant to help remove any temptation to think that such properties involve genuine relations at all. Once revised, the new statements refer only to *realia*—to real substances (subjects) with real accidents (thinking in a certain way).²⁷

To be sure, in some cases, there will be a self-standing *reale* that appears to “correspond” to what is thought by the judger. In others, however, such as in correct negative judgments, there obviously won’t “be” any further thing; the only *reale* involved will be the judger, who possesses a certain accident (thinking in a certain way). Brentano’s main point now is that, because what there is in *every* case is the judger, her acts, and certain modifications of these acts, the universal doctrine of truth and falsity should be built out of the consideration of these materials alone. What it means to say that something (a judgment) is true, ultimately, is to say that someone (a judger) is judging correctly. If there is any *adaequatio* between an *intellectus* and a *res*, the *res* in question is another *intellectus*—namely, that possessed by the correct *judger*.

Of course, Brentano now owes us an explanation of what it is to judge “with evidence,” what it is for a judgment to be “evidently correct,” to be the sort of thing the correct judger does. In particular, he owes us an account of that in virtue of which something is, or can be, judged with such evidence. Ultimately, however, Brentano appears to *deny* that any further explanation can be given for the experience of the evident correctness of a judgment. All we can do, he thinks, is “consider a multiplicity of evident judgments and then contrast them with other judgments that lack this characteristic,” just as “when we make clear to ourselves what is red and what is not red” (TE 125, 143), so that we can finally “see,” as it were, what it is to be an evident judgment. In this way, Brentano therefore takes the “characteristic” or “mark” of being evident itself to be explanatorily *primitive*.²⁸

²⁶ Brentano’s commitment to this sort of redescription-project can be seen in the following equation that he introduces during this period: “there is something thought of [*ein Gedachtes*]” = “there is a thinker [*ein Denkendes*]” (TE 68; WE 79).

²⁷ For more on Brentano’s late theory of judgment, see Chisholm (1982); on the turn toward adverbialism, compare as well Moran (1996).

²⁸ Here there would then be clear affinity between the views of Brentano on evidence and the views of Brentano’s contemporary, G. E. Moore, concerning our relation to such “simple” concepts as “good” and “yellow” (cf. Moore 1903: sections 7–10).

At this point, however, worries arise that Brentano has eliminated truth from view altogether. We might worry, more specifically, whether such a position runs the risk of sacrificing the idea of the objectivity of truth, of truth's being ultimately object- or fact-directed, its being ultimately beholden to the facts involving the objects represented. For Brentano's late analysis not only prohibits things like facts from playing any interesting role in an account of truth, but ultimately denies that there "are" any such facts, in "the proper sense" of "being." To the contrary, Brentano now wishes to take our primitive *experience* of evidence—experience of a state that we are in, or of a property that we possess, as being correct—as something that provides not just the entry-point for the analysis of truth, but rather a sufficient basis for completely reconstructing its content.²⁹

It is hard to see how primitive experiences of a sense of correctness *alone* could do full justice to what motivates all of the familiar appeals to more semantical considerations having to do with successful reference to an object of an intentional act.³⁰ For with its adverbialist construal of representational content, Brentano's late account threatens to detach truth entirely from its *relation* to those objects toward which we otherwise take ourselves to be engaged with, in and through our representations.

4.5 HUSSERL AND THE RETURN OF THE FACTS

Husserl, for one, came to think that his erstwhile teacher had gone too far at this point. Though Husserl agreed that any account of truth must be intimately linked to our experiences of evident correctness, he also thought that any plausible account of evidence itself simply must recognize a link in the other direction, between truth and the objectivities that were represented in the bearers of truth. In effect, though the connection between truth and intentionality-involving acts by subjects should not be severed along Bolzanian lines, its preservation should not come at the expense of the connection to the facts represented, which Brentano's account had threatened to obscure. For the latter to be achieved, Husserl saw no better hope than to return to Brentano's middle-period intuitions concerning the need for objective correlates of judgment—despite the challenges posed by Brentano himself.

By the late 1890s, however, it was Bolzano's influence—and in particular, Bolzano's way of drawing the distinction between mental act, logical content ("matter"), and metaphysical object—that pushed Husserl's thought decisively away from Brentano's. The embrace of this threefold distinction provided Husserl with an important platform from which to address Brentano's worry about the bloat in ontology that acceptance of states of affairs would cause,

²⁹ In this respect, Brentano's late account of truth might well be called "epistemic"; compare Parsons (2004).

³⁰ This is more than a little ironic, given the early Brentano's fame for having reintroduced intentionality as a central topic in philosophy.

since the reasons that support Bolzano's way of drawing the distinction between acts and contents are reasons that Husserl takes to *already* support an ontology that goes well beyond the bounds of Brentano's "reism."

When Brentano introduced the distinction between act and content, his account of this distinction did not push much further than an appeal to the testimony of an inner experience of a difference between, for example, the hearing of a sound and the sound as what is heard (cf. PES 78–80; I.111–12). Yet by describing the content as "in-existing" or as being "immanent" or "contained" within the act (PES 88; I.124–5), Brentano's language made it sound as if the content was connected to the act by being a real part of it. His later adverbialism about contents connects them even more explicitly to the reality of the acts of which they are modifications. Husserl, by contrast, joins Bolzano in taking the fact that several, distinct, real mental acts can nevertheless contain *identical* contents or meanings to show that there is a reason for drawing a deeper ontological distinction between acts and contents—namely, because these contents *cannot* themselves be real parts of acts or any other sort of real entity. While everything that is real, whether physical or mental, is individual and individuated by its absolutely singular place in time (cf. LU II section 8, I.249; II.123), "the essence of meaning," by contrast, lies in "one *identical* [*identische*] intentional unity set over and against the dispersed multiplicity of actual and possible experiences of speaking and thinking" (LU I section 30, I.228; II.97; emphasis added; cf. LU I section 31, I.229; II.99). We see this, thinks Husserl, when we realize, for example, that we cannot "confuse the 'judgment' as *content* of judgment, i.e. as an *ideal* unity, with the individual *real act* of judgment," since otherwise we could not speak of "*the* judgment '2 x 2 = 4'" as being "the same whoever makes it" (LU Prol. section 36, I.80; I.119; emphasis added).

Husserl, therefore, agrees with Bolzano that *any* plausible theory of knowledge ("doctrine of science") will have to acknowledge the ideal sameness of meaning across multiple mental acts—at the very least, if it itself is to be communicated to others or even to one's future self. A commitment to the ideality of contents (meanings; Bolzanian propositions) is required by the very idea of a theory as such, as a collection of communicable truths (as a species of propositions). Any theory which "denied" this ideal, objective semantical aspect (such as Brentano's) would "nullify itself [*hebt sich auf*]" and become "countersensical [*widersinnig*]" (LU Prol. section 32, I.76; I.112). From this Husserl takes it to follow, not only that at least *some* "ideal objects" have their own "proper justification alongside of individual (or real) objects," but even more so that "idealism" in *this* (Bolzanian, logical) sense "presents the *sole* possibility of a self-consistent theory of knowledge" (LU II, I.238; II.107; emphasis added).³¹

If Bolzano and Husserl are both right to think that reference to non-real, ideal entities is ineliminable for any coherent account of knowledge, then Brentano's ontological scruples begin to look misplaced. Of course, accepting these particular ideal entities (i.e. contents) need not necessarily entail that states of affairs (the being of A, the being-F of

³¹ This lies behind Husserl's claim that "logic as a science *must* be constructed out of Bolzano's work" (LU Prol. section 61 App, I.143; I.227).

A), or any other putatively necessary ideality, must be accepted as well. Even so, such entities cannot be rejected simply out of concern for ontological austerity.³² And if, as Brentano himself had at one point conceded, the appeal to such entities enables a much more satisfying account of the nature of truth—and in particular, the nature of the truth-making relation—then they should merit a second look.

Yet while this might help assuage the ontological worry, Husserl still owes Brentano an answer to his second epistemological worry about the introduction of these sorts of entities, concerning the seeming impossibility of explaining how we could ever become acquainted with them in the first place. Husserl's ultimate answer is that we *are*, in fact, "seeing" or "intuiting" whole states of affairs when we see the truth of a judgment: "If an act is knowledge [*Erkenntnis*] in the strictest sense, i.e. if we judge with evidence, then the objectivity is *given*. The state of affairs now stands before us, not merely putatively [*vermeintlich*], but rather actually [*wirklich*] before our eyes . . ." (LU Prol. section 62, I.145; I.230). But what can this mean, that a state of affairs, like the being on the table of the knife, is "given" or present before the mind?

Husserl's attempt to go beyond Brentano here rests upon the assumption that once we are in possession of the simple, direct Kantian intuition (sense-perception) of a real individual, we can then go on to "see" further things about it:

Every direct act of perception . . . can function as a grounding act [*Grundact*] for new acts . . . acts that, in their new manner of consciousness, at once bring about a new consciousness of objectivity, one which originally and essentially presupposes the grounding act. (LU VI section 46, II.282; II.618)

What Husserl has in mind here includes our ability to attend to the concrete parts of the knife (its handle, its blade) as its parts, but also our ability to attend to its more "abstract" parts, like the relation that it bears to the table—and, ultimately, its being on the table.

Now, Husserl accepts that something akin to intellectual activity is required to get ourselves in a position to do such seeing. Nevertheless, he wants to insist, first, that the resulting ways of relating to the given object are not *mere* intendings or thinkings about aspects of the object (as when we merely *wonder* whether the knife is on the table), since what is being meant in these cases is, in a clear sense, genuinely present before the mind: *I see* that the knife is on the table. What is more, what is subsequently uncovered or brought to light is still something about the same object, something that is true of it, something "grounded" in the object itself. Finally, Husserl rejects the idea that these preparatory acts are synthetic or combinatory acts in the sense supposed by the Kantian tradition. The new acts that are required to be able to "see" such *abstracta* as states of affairs are not acts that "make up" or "create" something that is not there at all, nor do they introduce things into our representations of things that find no match in what is ultimately given. Rather, the acts necessary for this new kind of intuition are more

³² Not unlike Frege, Husserl himself would add that Bolzanian idealism is motivated as well by considerations concerning the ontology necessary for arithmetic (cf. LU Prol. section 46).

appropriately described as a kind of *analysis* that lets us *abstract* from what is already sensibly present, by teasing out or lighting up for our consciousness the abstract, categorical structure of the object itself—an act that Husserl aptly calls “categorical abstraction” (LU VI, II.186; II.478).³³

At the same time, though, Husserl takes great pains to emphasize that it is not abstraction through *generalization*. He thinks that Brentano had been right to complain that construing the moves from S to the being of S, the being-P of S, etc., as moves from species to genus is nonsensical. Rather, Husserl takes the requisite act of abstraction to be one in which consists in a form of *partition*, one in which abstract “parts” or “moments” of the whole object are to be “set into relief [*heben heraus*]” and “new objects thereby emerge [*erwachsen sind*]” (LU VI section 48, II.286 and 289; II.624 and 628). As Husserl’s student Martin Heidegger would put the idea some years later, such acts “disclose [*erschliessen*]” or “lay open [*aufschliessen*]” the being (or being-P) of S (Heidegger 1927: section 16).

In this way, Husserl thinks that direct, sensible intuition is something that can go on to “found” or “ground” new acts of intuiting which *do* make present “new objectivities” that are constituted with just the right sort of (ontological) “categorical” structure (such as the being-P of S, etc.) to place them in correlation with (logical) propositional structure (such as the predicative structure: S is P). Hence, though it is true (as the late Brentano and Kant had both supposed) that “intentions containing categorical forms cannot find their fulfillment in *mere* sensibility,” such intentions *do*, nevertheless, have fulfillment in *something* that is “given” or “intuited”—namely, “in a sensibility that has been *formed* [*geformte*] through categorical acts” (LU VI, II.186; II.477; emphasis added). Husserl’s name for the exercise of this kind of “formed” sensibility is “categorical intuition” (LU VI, II.186; II.478).

On this picture, what the copula ultimately represents—and in true judgments, “corresponds” to—is not an *act* of mind at all (early Brentano), nor is it something “put” into our representations by an act (Kant), nor is it simply “mysteriously” there in the content without representing anything (Bolzano). Rather, just as the middle-period Brentano had intimated, the copula corresponds to something that is genuinely “out there,” something mind-independent and objective, something waiting to be “given” and so “fulfill” our judgments about objects:

Not in reflection upon acts of judgment . . . but in the fulfillments of judgments themselves lies the true origin of the concepts of states of affairs and of being (in the sense of the copula). Not these acts as objects, but in the *objects* of these acts do we have the foundation for abstraction which enables the realization of these concepts. (LU VI §44, II.279; II.613)

If this is right, however, then the analysis of the *truth* of propositions in terms of their genuine, full-fledged *agreement* with their objects is given new life. Like Bolzano and

³³ In this respect, Husserl is onto something not unlike the topic of Wittgenstein’s discussion of what is involved in “seeing aspects” of an object. For exploration of this analogy, see Bell (1990: 107f).

Kant, Husserl takes the primary significance of “true” (and “false”) to be the “logical” sense, of predicates that apply to contents—propositions—and not to the acts in which these are grasped or judged, nor to the objects (states of affairs) that such contents represent (LU Prol. section 47n, I.318n7; I.178). Yet unlike Bolzano or Kant, Husserl can now, finally, claim that what *makes* propositions true is the fact that there is an objective correlate that contains in it *nothing less than* what is represented in the proposition itself. And in its ability to fully reinstate, without qualification, the object-directness of the truth-relation, Husserl sees the notion of categorial intuition as providing, at long last, a genuine escape-route from the legacy of Kantian subjectivist idealism in the theorizing about the nature of truth.

The introduction of categorial intuition also had the further appeal for Husserl of allowing him to account for the experience of evidence or correctness in a way that gives the lie to the late Brentano’s claim that this phenomenon possesses absolutely self-sufficient or primitive intelligibility and is that in terms of which truth itself must be explained. Husserl agrees with Brentano that there is an intimate connection between our understanding of truth and our sense of the ideal possibility of having an experience of, “insight” into, the correctness of the corresponding judgment, in the sense of seeing it to be evidently correct: “The assertions ‘the truth obtains [*gilt*],’ and ‘there could be thinking beings who have insight [*einsehen*] into judgments with the relevant meaning-content’ are *equivalent* [*von gleichem Werte*]” (LU Prol. section 39, I.86; I.129; emphasis added).³⁴ Yet Husserl insists on supplementing Brentano’s account by continuing the circle of interconnections so as to show the notion of evidence itself to contain an intrinsic link to the intentional or representational relation to objects.

This is achieved, Husserl thinks, by recognizing, first, that the concept of being in general and the concept of an object in general are concepts that are intelligible only in correlation with the concept of an adequate perception of the relevant objectivity:

Perception and object are the most intimately interconnected concepts, concepts that reciprocally assign sense [*Sinn*] to one another. (LU VI section 43, II.277; II.610)

The equivalence of the concepts of being [*Sein*] and the possibility of adequate perception is undeniable. (LU Prol. section 50, I.118; I.185)

Evidence, as a kind of perception or experience, is no different in this basic respect, though its particular objectivity is of a special sort. This is because evidence is the “experience” of truth itself, which Husserl thinks we can now understand as the experience of the “fulfillment” of what is intended in one’s judgment (a proposition) by what is actually “given” in a categorial intuition (a state of affairs). Or, as Husserl also puts it, evidence, as

³⁴ Compare LU Prol. section 50, where Husserl states that a general “equivalence” obtains between the propositions “A is true” and “It is possible that anyone could judge with evidence that A is true” (I.117; I.184).

the experience of truth, is nothing other than the experience of the *adaequatio* between the *intellectus* and *res* (cf. LU VI section 37, II.261; II.590).³⁵

With this last stroke, then, Husserl provides us with a position in which all three of the dimensions we initially identified as belonging to the familiar conception of truth—the mind-independent fact-directness of truth (*res*), the felt sense of correctness or normativity we experience in relation to the truth (*adaequatio*), and, finally, the truth's link to our mind and its representations (*intellectus*)—are accorded a proper place, with none of these dimensions being reduced or eliminated or re-constructed in terms of another. Instead, Husserl proposes a relationship of mutual and necessary correlation, of “reciprocal sense-dependence,” between these notions.

For some, like the late Brentano, the ontological cost of this otherwise harmonious position will be too high, insofar as Husserl is asking us to accept that “there are,” in a genuinely objective sense, an infinity of infinitely complex *abstracta* waiting to be lifted out of the world of ordinary sense-experience. For others, perhaps like Hegel, Husserl's proposal of a mutual accord between the dimensions of truth ignores the necessity of a deeper inquiry into the reason or ground for such inter-correlations between concepts, the need to pursue questions like: Why do states of affairs have just the right structure to be apprehended as correlates of propositional contents? Why do propositions themselves have just the right structure to be grasped in judgments? Why are parts of reality able to be represented, “intended,” by what is ideal? And how are real acts able to grasp what is ideal?³⁶

4.6 CONCLUSION: THE RETURN OF THE SUBJECT?

Whether or not we find such questions well-motivated, it is worth noting, in closing, that, in the ensuing decade, Husserl himself was later moved to pursue an inquiry along just these lines. What is more, in his next major work, his 1913 *Ideas I*, Husserl argues explicitly that Kant was more or less on the right track after all, that the ultimate, “absolute” ground for such correlation does lie in what Husserl himself now calls “transcendental subjectivity.” Indeed, though it has been less common after Husserl to appeal

³⁵ Cf. LU Prol. section 51: “Evidence is nothing other than the ‘experience [*Erlebnis*]’ of truth . . . [T]hat which is evidently judged not merely judged (not merely meant in a judging, asserting, affirming manner) but is itself present [*gegenwärtig*] in the experience of judgment . . . The experience of the agreement [*Zusammenstimmung*] between the meaning [*Meinung*] and that which is present, experienced, as what it meant, between the experienced sense of the assertion and the experienced state of affairs—this is evidence” (I.121; I.190).

³⁶ These are, in fact, just the questions that the neo-Kantian Paul Natorp raises in his review of the *Prolegomena* to Husserl's *Investigations* (cf. Natorp 1901).

to “transcendental” grounds, many others throughout the twentieth century have followed suit in attempting to ground an account of truth in reflections on manifestations of human subjectivity, whether our mental activity, socio-linguistic practice, or our biological or neurological make-up. For those, by contrast, who, though perfectly willing to recognize a role for subjectivity to play in the analysis of truth, nevertheless want to retain an equal role for the objectivity that comes with fact-directness to play as well—for such traditionalists, Husserl’s early attempt to strike just the right, delicate, non-reductive, non-eliminative balance between the powerful intuitions of Kantian idealism, Bolzanian idealism, and Brentanian anti-idealism may provide something of more lasting value.*

REFERENCES

- Arnauld, A. and N. Pierre (1992) [1683]. *La logique ou l'art de penser*, 5th edn. Paris: Gallimard.
- Baumgarten, A. (1757). *Metaphysica*, 4th edn. Halle: C.H. Hemmerde.
- Bell, D. (1990). *Husserl*. London: Routledge.
- Benoist, J. (2006). “Variétés d’objectivisme sémantique,” in J. Benoist (ed.), *Propositions et états de choses*. Paris: Vrin, 13–49.
- Bolzano, B. (1837). *Wissenschaftslehre*. Sulzbach: Seidel. (“WL”)
- Bolzano, B. (1851). *Paradoxien des Unendlichen*, ed. F. Prihonsky. Leipzig: Reclam.
- Bolzano, B. (1935). *Der Briefwechsel B. Bolzano’s mit F. Exner*, ed. E. Winter. Prague: Royal Bohemian Society.
- Bolzano, B. (2004). *On the Mathematical Method and Correspondence with Exner*, trans. and ed. R. George and P. Rusnock. Amsterdam: Rodopi. (“CE”)
- Brentano, F. (1924) [1874]. *Psychologie vom empirischen Standpunkt*, 2 vols., ed. O. Kraus. 2nd edn. Hamburg: Meiner.
- Brentano, F. (1930). *Wahrheit und Evidenz*, ed. O. Kraus. Hamburg: Meiner.
- Brentano, F. (1933). *Kategorienlehre*, ed. A. Kastil: Meiner.
- Brentano, F. (1966). *The True and the Evident*, trans. and ed. R. Chisholm. London: Routledge. (“TE”)
- Brentano, F. (1973). *Psychology from an Empirical Standpoint*, trans. and ed. L. L. McAlister. London: Routledge. (“PES”)
- Brentano, F. (1981). *The Theory of Categories*, trans. and ed. R. Chisholm and N. Guterman. The Hague: M. Nijhoff. (“TC”)
- Chisholm, R. (1982). “Brentano’s Theory of Judgment,” in R. Chisholm, *Brentano and Meinong Studies*. Atlantic Highlands, NJ: Humanities.
- Descartes, R. (1898) [1639]. *Oeuvres de Descartes*, vol. 2, ed. C. Adam and P. Tannery. Paris: Leopold Cerf.
- Hegel, G. (1812). *Wissenschaft der Logik*, vol. 1. Nuernberg: Schrag.
- Heidegger, M. (1927). *Sein und Zeit*. Halle: Niemeyer.

* Many thanks to Michael Glanzberg, Ansten Klev, Samantha Matherne, Gila Sher, and Eric Watkins for comments on earlier drafts of this chapter.

- Henrich, D. (1997). "Hölderlin on Judgment and Being," (reprinted) in E. Förster (trans. and ed.), *The Course of Remembrance and Other Essays on Hölderlin*. Stanford, CA: Stanford University Press
- Husserl, E. (1900–1). *Logische Untersuchungen*, 2 vols. Halle: Max Niemeyer. [LU]
- Husserl, E. (2001). *Logical Investigations*, 2 vols., trans. and ed. J. N. Findlay. London: Routledge.
- Kant, I. (1902–). *Kants gesammelte Schriften*. Deutsche Akademie der Wissenschaft. Berlin: De Gruyter.
- Kant, I. (1991–). *Cambridge Edition of the Works of Immanuel Kant*. ed. P. Guyer and A. Wood. Cambridge: Cambridge University Press.
- Kemp Smith, N. (1918). *A Commentary on Kant's Critique of Pure Reason*. London: Macmillan.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Leibniz, G. (1978) [1705]. *Nouveaux Essais sur l'entendement humain*, in C. Gerhardt (ed.), *Die philosophischen Schriften*, vol 5. Hildesheim: Olms.
- Locke, J. (1979) [1689]. *An Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Oxford University Press.
- Martin, W. (2006). *Theories of Judgment*. Cambridge: Cambridge University Press.
- Marty, A. (1884). "Über subjectlose Sätze," *Vierteljahrsschrift für wissenschaftliche Philosophie*, vol. 8.
- Meier, G. (1752). *Auszug aus der Vernunftlehre*. Halle: Gebauer.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- Moran, D. (1996). "Brentano's Thesis," *Proceedings of the Aristotelian Society Supplementary Volumes* 70: 1–27.
- Natorp, P. (1901). "Zur Frage der logischen Methode. Mit Beziehung auf Edm. Husserls 'Prolegomena zur reinen Logik,'" *Kant-Studien* 6: 270–83.
- Parsons, C. (2004). "Brentano on Judgment and Truth," in D. Jaquette (ed.), *Cambridge Companion to Brentano*. Cambridge: Cambridge University Press, 185–93.
- Prauss, G. (1969). "Zum Wahrheitsproblem bei Kant," *Kant-Studien* 60: 166–82.
- Simons, P. (1987). "Brentano's Reform of Logic," *Topoi* 6: 25–38.
- Simons, P. (2004). "Judging Correctly: Brentano and the Reform of Elementary Logic," in D. Jaquette (ed.), *Cambridge Companion to Brentano*. Cambridge: Cambridge University Press, 45–65.
- Smith, B. (1989). "Logic and the Sachverhalt," *The Monist* 72: 52–69.
- Spinoza, B. (1677). *Ethica*. In *Opera Posthuma*. Amsterdam: Rieuwertsz.
- Van Cleve, J. (1999). *Problems from Kant*. Oxford: Oxford University Press.
- Walker, R. (1989). *A Coherence Theory of Truth*. London: Routledge.
- Windelband, W. (1884). *Präludien*. Tübingen: Mohr.
- Wolff, C. (1740). *Philosophia rationalis sive logica*. Leipzig: Francofurti.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

PART II

TRUTH IN EARLY
ANALYTIC
PHILOSOPHY

CHAPTER 5

TRUTH IN BRITISH IDEALISM AND ITS ANALYTIC CRITICS

THOMAS BALDWIN

5.1 INTRODUCTION

THE concept of truth was central to the debates between British idealists such as F. H. Bradley (1864–1924) and their critics such as Bertrand Russell (1872–1970). Bradley held that in discursive thought a thought's content is attributed to reality in such a way that the complete truth of the thought is accomplished only where the thought is expanded into a holistic mode of experience which encompasses the totality of reality. By contrast Russell, initially at least, took it that truth and falsehood are simple properties of propositions, which are complex things that we typically encounter as objects of thought but whose being is not dependent on thought and which constitute facts where they are true. As these brief descriptions indicate, the contrast here runs very deep: it is not just one of a clash of philosophical doctrines, idealist monism versus realist pluralism. Beyond this clash of doctrines, and despite the mutual respect in which Bradley and Russell held each other, there are two very different kinds of philosophical discussion and argument: where the idealists attempt to develop a synoptic metaphysics that, in Bradley's phrase, "satisfies the intellect" (1897: 554), their critics offer brisk, sharp-edged, arguments to show that the idealist philosophy rests on assumptions for which its proponents provide no arguments and whose implications are unacceptable. Russell famously described his conversion in 1898, thanks to the influence of his younger friend G. E. Moore (1873–1958), from the idealist philosophy to which he had initially adhered as "a great liberation, as if I had escaped from a hot-house on to a wind-swept headland" (1995: 48). But Bradley still ended his 1914 *Essays on Truth and Reality*, which is in large part a response to Russell, by affirming his commitment to exposing the shallowness of the "abstract" distinctions on which Russell relied: "And one main work of philosophy is to show that, where there is isolation and abstraction, there is everywhere, so far as this abstraction forgets itself, unreality and error" (1914: 473).

The central place of the concept of truth in these debates comes from the conception of logic that had developed in the work of the British idealists. “Logic” of a kind had had a central place in Hegel’s writings, but Hegel’s logic was primarily concerned with questions about being, non-being, and becoming—what we might now call ontology. By contrast the logic texts produced by Bradley, Bernard Bosanquet (1848–1923) and others focus on topics much more familiar to us, such as judgment, negation, conditionals, and inference. These texts are not works of formal logic: they are very different from J. N. Keynes’s *Formal Logic* (1884), the last major work of logic in English within the syllogistic tradition, let alone George Boole’s *Laws of Thought* (1854) which initiated a radically new algebraic program in formal logic among logicians such as De Morgan, Venn, and Jevons. Instead the logic texts of the British idealists are works of what we might now call “philosophical logic” or “theory of meaning.” Bosanquet briefly sums up the fundamental question of logic in this way: “How does the course of my private ideas and feelings contain in it, for me, a world of things and persons which are *not merely in my mind?*” (1895: 4; emphasis in original). And it is not only in their logic texts that questions of this kind are given attention by the British idealists; Bradley devotes chapter 15 on “Thought and Reality” of *Appearance and Reality* to his account of thought, truth, and reality, and he later remarks that this chapter “contains the main thesis of this work” (1897: 555). So in this respect, at least, the British idealists anticipated the “revolution” in philosophy characteristic of their analytic critics by basing their metaphysics on their theory of meaning.

It is, therefore, not surprising that questions about truth are prominent both in the works of the British idealists and in those of their critics. One mark of this is that it is in the debates to which they gave rise that the familiar “theories” of truth are first regularly discussed as such, especially the “correspondence theory” and the “coherence theory” (see e.g. Moore 1902). Equally it is in this context that “the nature of truth” becomes for the first time a topic for sustained discussion, most notably in Harold Joachim’s book, *The Nature of Truth* (1906). For while there are of course many earlier works on “truth” (such as Edward Herbert’s *De veritate* (1624)), these works typically make claims concerning fundamental truths and our methods of identifying them. By contrast Joachim’s book, which is discussed below, makes few claims of this kind; instead his main purpose is to argue against correspondence theories of truth, and also Russell’s thesis that truth is a simple quality of propositions, and to defend in their place a coherence theory of truth.

5.2 RELATIONS

An issue that is intertwined with the debates about truth is that of the status of relations. Bradley held that discursive thought of the kind which constitutes knowledge centrally concerns relations between terms; but also that all relations are “unreal.” Both these claims need considerable unpacking and I shall return to them, but their connections

with truth can already be briefly indicated. Given the unreality of relations, it follows that the truth of thoughts cannot be either an external correspondence with a relational fact or, indeed, the reality of the thought's relational content. Instead, the unreality of relations implies that apparent relations between terms must be grounded in some non-relational system, and thus that the apparent plurality of different, but related, terms is grounded in their just being aspects of one all-encompassing system, the Absolute. So truth admits of degrees, in that as thoughts are refined and connected in the light of the systematic connectedness of their apparently relational contents, the thoughts become more true. But the complete truth of thought requires its identity with reality, which, however, transforms thought into a new non-relational and non-discursive mode of experience.

Russell agreed with Bradley that a proper understanding of relations is essential for logic and metaphysics. In his early work on the philosophy of mathematics, for example, he emphasized the importance of relations (see chapters 9, 26 of *The Principles of Mathematics* (Russell 1903)). But he rejected Bradley's thesis of the unreality of relations and the associated thesis that relations can therefore be at best only an apparent implication of the non-relational properties of some wider system. Russell had made his rejection of a thesis of this kind a central theme of his early book on Leibniz (Russell 1937 [1900]), though his attribution of this thesis to Leibniz is now generally accepted to have been a mistake (Ishiguro 1972). When discussing Bradley's "monistic theory of truth" Russell attributed the same thesis, which he called the "axiom of internal relations" (Russell 1910: 139), to Bradley and argued that Bradley fails to provide any satisfactory arguments in support of this "axiom" and that it leads to insuperable difficulties concerning the manifest complexity of the world (Russell 1910: 145–6). As we shall see below, Russell's account of Bradley's position, in particular the emphasis placed on "the axiom of internal relations," is questionable. Nonetheless it is clear that Russell's different treatment of relations, which culminates in his "multiple-relation" theory of judgment,¹ is associated with his very different accounts of truth, first as a simple property of certain propositions and then as a correspondence between judgments and facts.

Before spelling out the details of these complex debates, however, it is worthwhile going back to the treatment of relations propounded by T. H. Green (1836–82). Green was the first important British idealist philosopher, but whereas Bradley's idealism is broadly Hegelian, Green's idealism is rooted in his study of Kant's critical philosophy, combined in his case with familiarity with the works of the British empiricists, notably Locke and Hume. Indeed Green was responsible for the first complete edition of Hume's writings, though his lengthy introduction to the edition largely concerns Locke's theory of ideas. It is, then, though reflections on Locke's thesis that relations are "not contained in the real existence of things" (*Essay* II.25.8 (Locke 1975 [1689]: 322)) but "superinduced" by

¹ Some of the authors discussed in this chapter write "judgment" where others write "judgement." In all direct quotations I have preserved the spelling used by the author, but otherwise I have just used the spelling "judgment" myself.

our “comparing or considering two things together” (*Essay* II.25.7 (Locke 1975 [1689]: 322)) that Green is led to reformulate Kant’s thesis that “the understanding makes nature” in terms of relations (Green 1883: 15). Thus in his 1874 Logic lectures Green maintains that “all reality lies in relations” and that “only for a thinking consciousness do relations exist” (Green 1886: 177). Green explains that by the first point, that “all reality lies in relations,” he means that insofar as we take anything to be real, we take it to enter into relations, most notably space and time (Green 1886: 170). But, he goes on to insist—and this is the second point—these relations exist only as aspects of a world which makes sense as a unified system of nature for a consciousness. Green recognizes that this position will be criticized for not accommodating a conception of reality which is altogether external to consciousness, such as Kant’s things-in-themselves. But Green argues that there is no warrant for this conception, which in Kant’s case ends up undermining his own account of empirical reality (Green 1883: 40–3). However Green acknowledges that he needs to provide a basis for our having a conception of the world as one complete and unified system, as opposed to the many partial systems affirmed by different subjects. Hence he is led to the speculative hypothesis that the consciousness of empirical subjects is just a partial realization of one eternal self-distinguishing consciousness for which the real world exists, of which he writes:

It has no place in nature, except as determined by relations which can only exist for a thinking consciousness. For the consciousness which constitutes reality and makes the world one it exists, not in that separateness which belongs to it as an attribute of beings that think only at times or not at all, but as conditioned by a whole which thought in turn conditions.

As to what that consciousness in itself or in its completeness is, we can only make negative statements. *That* there is such a consciousness is implied in the existence of the world; but *what* it is we only know through its so far acting in us as to enable us, however partially and interruptedly, to have knowledge of a world or an intelligent experience. (Green 1883: 53–4; emphasis in original)

At this point Green’s speculative idealism seems to be heading in the direction of theology, though in his *Prolegomena to Ethics* Green is primarily concerned to put his conclusion to work in developing his Kantian account of moral freedom. However, turning back now to questions about truth, what little Green says on this topic points clearly in the direction of a coherence theory:

It is not that there is first the reality of things, and then a theory about it. The reality *is* a theory. No motion is properly a phenomenon, but a relation between phenomena constituted by a conceiving mind; a way of holding together phenomena in thought. (Green 1886: 269; emphasis in original)

So the truth of a theory is not a matter of its correspondence to the reality of thing; instead the reality just *is* the theory which captures the observed phenomena in

such a way that they “hold together,” i.e. cohere. As such, Green holds, all truths are necessary:

Given any proposition conceived as wholly (unconditionally) true, you cannot conceive its contradictory to be true consistently with that idea of the unity of the world without which no proposition could be conceived to be really either true or untrue. (Green 1886: 266)

5.3 BRADLEY AND THE DEMANDS OF TRUTH

After his early death in 1882, Green’s influence among the British idealists was increasingly supplanted by that of his younger Oxford colleague, F. H. Bradley. So far as I am aware, in his published writings Bradley never refers explicitly to Green, and there are no obvious implicit references either. However it is hard not to connect Bradley’s emphasis on relations, and in particular his characterization of discursive thought as essentially “relational” with Green’s emphasis on relations. But it is equally clear that Bradley’s treatment of relations differs sharply from Green’s—where Green affirmed that “reality lies in relations” Bradley holds that relations are “unreal,”—by which he does not just mean that they exist only for consciousness.

So what then does the “unreality” of relations amount to? According to Bradley, the test of reality is that it be “self-existent, substantial, and individual” (1922: 71). Applying this test to relations, Bradley argues, it is clear that they are not real. If we consider, first, a relation by itself, such as being older than, this is not a self-existent, substantial, individual. Indeed Bradley holds that a relation so conceived is “mere verbiage”—“a phrase without meaning” (1897: 32). If we add a pair of terms and obtain a relationship, such as Green’s being older than Bradley, however, this conclusion is not so immediately clear: is not this relationship at least self-existent and individual? At this point, however, Bradley introduces his famous “regress argument” to argue that no such relationship, at least as conceived of as obtained by combining a relation to two terms, is self-existent:

But here again we are hurried off into the eddy of a hopeless process, since we are forced to go on finding new relations without end. The links are united by a link, and this bond of union is a link which also has two ends; and these require each a fresh link to connect them with the old. (1897: 32–3)

This argument assumes that the relation within a relationship needs to be tied to its terms by a new, linking, relation; and given this assumption the argument is sound, and shows that no such relationship is self-existent. But Bradley’s critics will reject the assumption and argue that, properly understood, relationships are not obtained by linking a relation such as being older than to two terms, say Green and Bradley, but are simply

ways in which the terms themselves are connected: because the “link” is the relation itself, it does not need any further links. On this alternative view, while the relation “by itself” is an abstraction from particular relationships and fails Bradley’s test for reality, particular relationships pass the test.

There are two further aspects of Bradley’s discussion of this point that merit brief attention here. The first is that although Bradley puts all his emphasis on relations, he acknowledges that similar considerations apply to qualities: in themselves they are also unreal because they are not substantial, and an object’s having a quality is not self-existent because of the need for a further relation to tie the quality to the object. The second point concerns the “unity” of judgment. Bradley presents his regress argument as one concerning the reality of relations, and I interpreted it above as implying the unreality of relationships; but one can also use much the same line of argument to assess the reality of subject-predicate and relational judgments. For Bradley, the relevant test of reality is as ever whether judgment is “self-existent,” and in this context this implies that the unity of a judgment should not be dependent upon connections which are not internal to the judgment. It is then easy to see how for Bradley no such unity can be achieved by just bringing together terms and a relational concept (1922: 96), and this leads him to reject the traditional doctrine of the copula as the characteristic link that holds together subject and predicate—“mere superstition,” as he puts it (1922: 21). As before, Bradley’s critics will object that his argument fails to take account of further alternatives such as the Fregean conception of concepts as fundamentally “unsaturated,” and thus in no need of a copulative link to form a complete proposition with a name (see “Function and Concept,” reprinted in Frege 1984: 137–56). Frege’s position is of course not free from problems, concerning, for example, the status of the concept “horse”; but we need not pursue these here. For it is reasonable to hold that some version of the Fregean position provides a way of dealing with the issue of the unity of judgment, and thus that Bradley’s challenge to relational judgments can be met.

In this discussion I have separated two interpretations of Bradley’s thesis of the unreality of relations, an ontological one concerning the unity of relationships and a logical one concerning the unity of judgment. These two lines of thought are of course connected and Bradley does not always separate them in his discussions of relations. But it is nonetheless worth keeping them apart at least provisionally, as the shortcomings of Russell’s response to Bradley indicate. For the main theme of Russell’s criticism of Bradley, his commitment to what Russell took to be the objectionable “axiom of internal relations,” can be thought of as arising in the context of what Russell took to be Bradley’s way of dealing with the ontological unity of relationships. Yet Bradley’s discussion of truth occurs primarily in the context of his discussion of judgment, which Russell does not clearly confront, so that Russell’s criticisms of Bradley’s “monistic theory of truth” fail to address Bradley’s main reasons for his theory of truth.

Nonetheless, it is simplest to follow Russell’s lead and start with the issue of internal relations. The issue here is that of securing the unity of relationships. I suggested earlier that this issue is misconceived; but that line of thought should be set aside if one is to understand Bradley’s route here to the Absolute and his tangled dispute with Russell

concerning internal relations. It is easy, first, to understand how one might suppose that merely “external” relations cannot provide the links necessary to secure the unity of a relationship between two terms. Hence, Russell supposed, the Bradleian solution to this problem was to conceive of the relation as “internal” in the sense that it is grounded in the “nature” of the terms involved; for this grounding looks, *prima facie*, to provide a way of securing the required unity of the relationship. Equally, however, Russell argued that the resulting position is problematic, for it requires a conception of the “nature” of a term which includes all of the term’s properties and therefore implies that there can be no substantive question as to whether a term has some property, a conclusion which Russell implies is absurd (Russell 1910: 144–5).

Without disputing this argument, however, Bradley, rejected Russell’s account of his position:

As to what has been called the axiom of internal relations, I can only repeat that “internal” relations, though truer by far than “external”, are, in my opinion, not true in the end. (Bradley 1914: 312)

Bradley explains why internal relations will not suffice in his later essay “Relations,” in a passage in which he imagines himself debating the issue with Russell:

You, I presume, still in this case would continue to take the terms each one as, so far, in and by itself real, and as independent absolutely of any whole that could be said to contain them. (Bradley 1935: 642)

Bradley’s presumption here is, I think, correct. For while Russell’s axiom implies that a relationship such as Green’s being older than Bradley is “grounded” in the ages of each of them, where this is taken to be part of their “nature,” this nonetheless leaves the existence of each of them “independent absolutely of” their relationship. As such, one can see why Bradley remains dissatisfied; for the unity of this relationship between Green and Bradley is not secured merely by their different ages. Bradley therefore takes it that the unity of the relationship requires a surrender of this presumed “absolute independence of the whole,” so that in the case of Green’s being older than Bradley, Green and Bradley have to be conceived as abstracted aspects of a unified system which includes both of them along with their different ages. And once one has reached this point the ascent to the Absolute is straightforward: since everything is somehow related to other things, each of these relationships can exist only if its terms are conceived as abstracted aspects of a unified system which includes everything—the Absolute. And precisely because the Absolute accommodates all relationships as abstracted aspects of this one system, it is itself “supra-relational” (Bradley 1935: 649), as Bradley maintains, and thus does not fall foul in its turn of the unreality of all relations.

So far as I can see, once one has set aside the misconception which starts off this whole construction, Bradley’s argument here is sound, and he is also right to maintain that Russell’s axiom of internal relations does not fully capture his position, although it is an

implication of it. Bradley in fact makes much stronger claims; concerning two red-haired men, for example, he writes:

By being red-haired the two men are related really, and their relation is not merely external . . . The correlation of the other circumstances of and characters in the two men with the quality of red-hairedness cannot in other words possibly be bare chance. And if you could have a perfect relational knowledge of the world, you could go on from the nature of red-hairedness to those other characters which qualify it, and you could from the nature of red-hairedness reconstruct all the red-haired men. In such perfect knowledge you could start internally from any one character in the Universe, and you could from that pass to the rest. (Bradley 1897: 580)

The assumption Bradley is bringing in here is that the only system which can sustain even mundane relationships, such as having the same colored hair, is one which determines all the “circumstances and characters” of the related terms since all these circumstances and characters also ground relationships, so that they too have to be encompassed in the system in which the red-haired men appear as just abstracted aspects. The resulting position is reminiscent of Green’s thesis, mentioned above, that all truths are necessary. But its implication that “you could start internally from any one character in the Universe, and you could from that pass to the rest” is very hard to accept, even harder than the axiom of internal relations which Russell refused to swallow.

I turn now to the logical interpretation of the unreality of relations, concerning the unity of judgment. Bradley in fact first deploys his regress argument in this context (Bradley 1922: 96) to undermine the conception of judgment as a “synthesis of ideas.” The positive account of judgment which Bradley contrasts with this conception is that “Judgement proper is the act which refers an ideal content (recognized as such) to a reality beyond the act,” where “The ideal content is the logical idea, the meaning as just defined” (Bradley 1922: 10). This account of judgment then enables him to present a preliminary account of truth: “By the truth of a judgement we mean that its suggestion is more than an idea, that it is fact or in fact” (ibid.).

Two quick points can now be made. First, the fact that judgment essentially involves a “reference to” reality which, if successful, implies its truth indicates that for Bradley what is fundamental to judgment is precisely this susceptibility to truth and falsehood (Bradley 1922: 2). And this suggests that instead of thinking of judgments as constructed from constituent ideas that are independently defined, one should think of ideas as abstracted from judgments which by “referring” their ideal content to reality affirm its truth. Bradley does not lay out this point as clearly as, say, Frege does with his context-principle (“Introduction,” *Foundations of Arithmetic* (Frege 1953: x)); but it is, I think, implicit in his treatment of judgment. It is certainly explicit in Bosanquet’s Bradleian introduction to logic, *The Essentials of Logic*:

The name or concept has no reality in living language or living thought, except when referred to its place in a proposition or judgement. We ought not to think of

propositions as built up by putting names or words together, but of words or names as distinguished though not separable elements in propositions. (Bosanquet 1895: 87)

(The term “proposition” is used here by Bosanquet to mean “A Judgement expressed in words” (1895: 82); this differs from its use by Moore and Russell discussed below.) The second point to make briefly here is that it will be clear that Bradley’s account of truth looks challenging in the light of the previous discussion of the ontological aspect of the unreality of relations. For if the truth of the judgment that Green was older than Bradley amounts to its being a fact that Green was older than Bradley, then we have to ask how there can be such a fact without its bringing in the whole Absolute system of which this fact is only one abstracted aspect, even though the system was no obvious part of the original judgment’s content. We shall see that dealing with this difficulty turns out to be central to Bradley’s discussion of judgment and truth.

Bradley begins *The Principles of Logic* (1922) by rejecting psychologistic accounts of ideas and arguing instead for a conception of ideas as meanings (1922: 6) which specify how things might be without entailing their existence; as Bradley puts it, ideas presuppose an abstract separation of content (“what” things are) from existence (“that” they are) (Bradley 1897: 162–3). This “separation” of content from existence is essential to thought of all kinds; but it is characteristic of judgment that it aims to overcome this separation by referring, or reattaching, the ideal content to reality: “Judgement is essentially the re-union of two sides, ‘what’ and ‘that’, provisionally estranged. But it is the alienation of these aspects in which thought’s ideality consists” (Bradley 1897: 165). Truth, then, is the successful achievement of this reunion—where the reference of ideal content to reality succeeds because it is a fact. Taken just by itself this suggestion admits of a modest interpretation as a “factive” variant of the deflationary account of truth:

The judgment that Green is older than Bradley is true iff it is a fact that Green is older than Bradley

I think this is indeed a legitimate way of expressing Bradley’s position, except that the word “deflationary” is not appropriate once one takes into account the weight that Bradley attaches to the existence of a fact. For the ontological interpretation of the unreality of relations in effect inflates facts into the Absolute. Bradley puts the point this way:

Truth is the predication of such content as, when predicated, is harmonious, and removes inconsistency and with it unrest. And because the given reality is never consistent, thought is compelled to take the road of indefinite expansion. (Bradley 1897: 165)

The important assumption here is that “partial truths,” such as that Green was older than Bradley, are not absolutely true since the fact that Green was older than Bradley is not a “self-existent individual.” Hence, in pursuit of its internal norm of truth, “thought is compelled to take the road of indefinite expansion”; the judgment that Green was

older than Bradley has to be inflated to encompass everything, since only where this is achieved can judgment's ideal content be the one all-encompassing fact—the Absolute. What is then distinctive of Bradley's position is that this achievement brings with it the transformation of thought; for once thought successfully encompasses everything, the separation of ideal content from existence that was essential to thought has been overcome. I shall return below to this striking conclusion, which Bradley celebrates as thought's "happy suicide" (1897: 173). But first I need to attend to the implications of this discussion for the attribution to Bradley of further conceptions of truth beyond the factive schema discussed above.

Bradley's reference to "the road of indefinite expansion" that thought is compelled to take in the pursuit of truth is suggestive of a coherence account of truth. For it looks as though this implies that truth is attained where, by means of indefinite expansion of belief, thought provides a fully coherent, or comprehensive and coherent (as Bradley often puts it (Bradley 1914: 219)), web of beliefs which encompass the totality of experience of all kinds. Yet although Bradley does say that coherence and comprehensiveness are aspects of "system" which is an "arbiter of fact" (1914: 219), and that as judgments become more systematic they become "truer and more real" (Bradley 1922: 639), he does not maintain that coherence and comprehensiveness, or system, are constitutive of truth. Instead Bradley's repeated claim is that the truth of a judgment consists in the unity of its ideal content with real existence. Hence, as he puts the point emphatically: "we must accept once for all the identification of truth with reality" (Bradley 1914: 113). It is, for Bradley, constitutive of the truth of judgment that its ideal content be real; and once, to meet this demand, "the road of indefinite expansion" has been traveled, the truth of judgment is identical with its reality. So Bradley's factive schema for truth should be interpreted as implying a conception of truth as identity with reality.

In *The Principles of Logic* (Book I, chapter 2) Bradley explores a rather different line of thought concerning truth. This line of thought starts from a discussion of a problem faced by the "copy" or correspondence theory of truth when confronted by disjunctive and hypothetical judgments—namely whether there are disjunctive or hypothetical facts to which such judgments correspond when true (Bradley 1922: 46). Bradley takes it as obvious that there are no such facts, and this leads him to ask whether there are any types of judgment which less problematically correspond to a fact or, to put the matter his way, whose truth directly expresses a fact, and the suggestion which he explores is that "Where the subject, of which you affirm categorically, is one individual, or a set of individuals, your truth expresses fact" (1922: 48). What is then striking is that in exploring this suggestion Bradley does not immediately introduce his thesis of the unreality of relations to shoot it down. Instead he enters into an extended discussion of singular judgment in which he deals with names and definite descriptions before settling on "analytic judgements of sense" such as "This is a tree" as judgments which, on the face of it, are indeed apt to express a fact when true. His reason for this conclusion is that the use of the indexical "This" makes it possible to fix a determinate fact that is expressed by the judgment where it is true because of the involvement of the actual context of the judgment given in perception in determining this fact. Two aspects of the use of indexicals

are important here: first, this use brings with it the successful reference to reality that Bradley takes to be a requirement of truth; second, because the reality referred to is provided by the actual context of the judgment, and does not simply occur in the ideal content of the judgment (as when I unhelpfully say “I am here” in answer to the question “Where are you?”), the particular fact expressed is determinate. Bradley puts this second point as follows:

We escape from ideas, and from mere universals, by a reference to the real which appears in perception. It is thus our assertion attains the uniqueness without which it would not correspond to the fact. And analytic judgements, it may seem, are thus secured to us. (Bradley 1922: 69–70)

It is worth emphasizing at this point that the status that Bradley assigns here to these indexical perceptual judgments does not arise from a foundationalist epistemology; indeed Bradley explicitly disavows any such epistemology (Bradley 1914: 290–309). Instead, their status is that they are uniquely apt to “correspond to the fact” (as he puts it here), and thus to act as the basis for the provision of indirect truth-conditions for other types of judgment. Thus the position Bradley sets out here resembles logical atomism, with these “analytic judgements of sense” having the role of atomic propositions. And yet, no sooner has Bradley set out this position, than he withdraws it: “analytic judgements of sense are all false,” he affirms (Bradley 1922: 93). Why? The reason he gives is that no judgment that is absolutely true can be such that its truth is context-dependent. For truth requires that the content of judgment be a fact, be real; and reality is, as ever, “self-existent, substantial, and individual.” Hence the context-dependence of the truth of these indexical judgments is incompatible with their being “really” true;

The analytic judgement is not true *per se*. It can not stand by itself . . . What it says is true, if true at all, because of something else. The fact it states is really fact only in relation to the rest of the context, and only because of the rest of that context. It is not true except under that condition. So we have a judgement which is really conditioned, and which is false if you take it as categorical. To make it both categorical and true, you must get the condition inside the judgement. You must take up the given as it really appears, without omission, unaltered, and un-mutilated. And this is impossible. (Bradley 1922: 97–8; cf. 1897: 179)

It is hard not to be disappointed by this conclusion. While Bradley is of course quite right to observe that one cannot bring the context of an indexical into the judgment so as to fix its reference unequivocally from within, what one would normally take to follow is that context-dependent indexicals have an irreducible and fundamental role in grounding the truth-conditions of language—which is precisely the conclusion which Bradley appeared to be affirming in the previous passage. Indeed Bradley’s discussion of this topic, with his emphasis on the role of the perceptual context in fixing the truth-conditions of indexical judgments, has insights which connect with more recent work on the subject. But all this detail is swept to one side in the face of his metaphysical

juggernaut, the demand that absolute truths must be self-existent, and thus cannot be in any way dependent, including context-dependent. Just as the all-embracing modal holism to which his ontological arguments lead calls into question the key assumptions of that argument, in this case, likewise, it seems sensible to hold on to context-dependent truth and reject his conception of the demands of absolute truth.

At this point therefore the discussion of “analytic judgements of sense” has led back to the conclusion we have already encountered, that judgments can never be true. In reflecting on this conclusion, however, we should note that Bradley has provided two routes to it and two ways of elaborating it. The route which we have just explored leads to the sceptical conclusion that, strictly speaking, all judgments are false, since absolute truth has to be unconditional, and there are no unconditional truths. The other route is not so unequivocally sceptical: this is the route whereby the truth of any one judgment requires the “indefinite expansion” of judgment to include all truths. For while this demand is not humanly achievable, it is not as incoherent as the demand that an indexical judgment should include its own context while still preserving its determinate truth-condition. Nonetheless, because the achievement of absolute truth by indefinite expansion would overcome the separation of ideal content from real existence which is essential to thought and judgment, according to Bradley this reunion of content with existence would transform judgment into a different mode of experience. This different mode of experience would still be true, but it would not be true judgment. Here is Bradley’s description of this transformation, in which true apprehension has become “quite identical with reality”:

Let us try to realize more distinctly what this supposed consummation would involve. Since both truth and fact are to be there, nothing must be lost, and in the Absolute we must keep every item in our experience . . . But to reach a mode of apprehension, which is quite identical with reality, surely predicate and subject, and subject and object, and in short the whole relational form, must be merged . . . Such a whole state would possess in a superior form that immediacy which we find (more or less) in feeling; and in this all divisions would be healed up . . . Every flame of passion, chaste and carnal, would still burn in the Absolute unquenched and unabridged, a note absorbed in the harmony of its higher bliss. (1897: 172).

5.4 JOACHIM AND COHERENCE

While Bradley affirmed that coherence is a criterion of truth but not constitutive of it, his younger friend and colleague Harold Joachim (1868–1938) advanced the thesis that “systematic coherence” is constitutive of truth in his book, *The Nature of Truth* (1906). Joachim begins with a critical discussion of the correspondence theory and also of Russell’s account of truth as a simple property of propositions which is described below. The central theme of Joachim’s criticisms of these theories is his denial that there are

mind-independent facts for our beliefs to correspond to: “the ‘facts,’ to which my judgement is to correspond, are the expression in and through my mind of that which is also expressed in and through the minds of my fellows” (1906: 23). So “the notion of correspondence must give place to the idea of systematic coherence” (1906: 29)—the systematic coherence of the common understanding which fixes the facts. In this respect Joachim’s position looks back to Green’s view that the relations that are constitutive of reality (of “facts”) are themselves the work of the mind; and the task which Joachim then undertakes is one of developing and defending this account of truth as systematic coherence.

But one thing that is attractive about Joachim’s discussion is that he avoids appealing directly either to Green’s mysterious “eternal self-distinguishing consciousness” or to Bradley’s equally mysterious supra-relational Absolute. Instead in order to convince us of the role of that “Ideal Experience,” as he calls it (1906: 83), whose complete coherence constitutes truth, Joachim elucidates our practices both in the natural sciences and in historical inquiries to show how our particular beliefs, “judgements of fact,” draw on an “inarticulate background” which, once made explicit, confirms our beliefs but also qualifies and corrects them:

What [the judgment of fact] affirms is subject to a complex mass of conditions unexpressed and yet implied. It draws its meaning and its truth from an inarticulate background of this kind. The judgement of fact, indeed, if it is to affirm *definite* meaning, demands the articulate expression of this background in the form of an explicit system of judgements. And yet in that system the original judgement, as formulated in isolation and as the mere statement of fact, would no longer persist. (1906: 107; emphasis in original)

This conception of the “inarticulate background” of common-sense belief anticipates many themes of twentieth-century epistemology; but precisely because the line of thought here seems to be primarily a contribution to epistemology, it gives rise to the anxiety that Joachim’s coherence theory is, in the end, a theory about the grounds for knowledge and not an account of truth itself. Joachim shows himself to be aware of this objection, that

we have failed to show that the “truth” of human knowledge is a “symptom” of the ideal truth, which we described in terms of coherence. We have failed to show that the substantial significant whole expresses itself both in our knowledge and in the reality known . . . (1906: 117)

To meet this objection Joachim embarks on a rash dialectical strategy. In order to show how the apparent contrast between the knowing mind and the reality known can be overcome, he seeks to show how the possibility of error can be accommodated within his coherence theory, since error appears, on the face of it, to involve precisely the kind of mind/world dualism which underpins the objection to conceiving of coherence as

constitutive of truth as well as essential to knowledge—“[A theory of truth as coherence] must show e.g. how the complete coherence, which is perfect truth, involves as a necessary ‘moment’ in its self-maintenance the self-assertion of the finite modal minds: a self-assertion, which in its extreme form is Error” (1906: 170). But having set himself this task Joachim has to admit that he cannot fulfill it: “the demands thus made cannot be *completely* satisfied by any metaphysical theory” (1906: 171; emphasis in original). The reason he gives is that only a theory which transcends the limitations of finite, human minds could achieve this goal, though quite why this objection is so decisive is not well explained: for even though human fallibility is incurable, it does not imply that there can never be a satisfactory human explanation of human fallibility. However that may be, Joachim takes it that this failure is critical: “the coherence-notion of truth may thus be said to suffer shipwreck at the very entrance of the harbour” (1906: 171); and he infers from this “wreck” that the coherence theory of truth is self-undermining:

Assuming that the coherence-notion of truth is sound, no theory of truth as coherence can itself be completely true, but is at most possessed of a “truth” which we may believe, but have not proved, to be “symptomatic” of perfect truth. (Joachim 1906: 175)

This final collapse of Joachim’s attempt at a defense of the coherence theory of truth tends to confirm the thesis that although coherence has a fundamental role in epistemology, it is not a sound basis for a theory of truth.

Russell, who was on friendly terms with Joachim, reviewed Joachim’s book, but concentrated in his review on Joachim’s thesis that “facts” are not mind-independent (Russell 1906). Russell takes it that, for Joachim, this thesis depends on the rejection of external relations, including the conception of experience as a relation of this kind. Although this provides only a limited account of Joachim’s line of thought (which, as I indicated, strikes me in this respect as closer to that of Green than, as Russell assumes, that of Bradley), it is fair enough as far as it goes. What is then striking is that Russell here suggests that once the disagreement is taken to be focused round this issue of whether there can be external relations, constructive argument comes to an end:

The difficulty, as regards the second question [whether any arguments are possible which ought to appeal to both sides], lies in the fact that our differences are so fundamental that almost all arguments on either side necessarily begin by assuming something which the other side denies. (Russell 1906: 528)

So at this point Russell seems to accept that the disagreement between idealists such as Joachim and Bradley, on one side, and their critics, such as himself and Moore, go so deep that there is no way to take the debate forward between them. As we saw above, Russell did also think that the implications of the “axiom of internal relations” are unacceptable, but he would also have acknowledged that these implications were not in fact unacceptable to Joachim and Bradley.

This question, of whether there was any way of taking forward this debate, was then taken up by Moore and Joachim. For Moore followed up Russell's review with a discussion note in which he suggested that there is a way of developing an objection to Joachim's position which cannot be easily set aside, namely by identifying an implication of the position which Joachim could not easily accept. In this case, Moore suggests, Joachim's rejection of external relations implies that the possibility of disagreements, such as that between himself and Joachim, is ruled out; for Joachim's position implies that it cannot be that one and the same proposition is both affirmed by one person, Joachim, and denied by another, Moore. And yet, Moore observes, "I find it hard to believe that Mr. Joachim at Oxford is totally incapable of denying precisely and numerically the same truth, which I affirm at Edinburgh" (Moore 1907: 235). This clever argument is characteristic of one of Moore's approaches to philosophical dispute, that of finding a pragmatic incoherence in the position of his opponent (Moore famously uses it in his discussions of skepticism). Unfortunately Joachim, in his response to Moore (Joachim 1907), does not really engage with Moore's argument. He insists that the judgments of people in different situations are affected by their situations, so that their judgments do indeed differ—which appears to rule out the possibility of disagreement. But he then goes on to allow that judgments can sometimes be "abstracted" from their situation, in such a way that, it would seem, we can make some sense of the occurrence of disagreements (though Joachim does not say so). In the end it is hard to know quite what response to Moore's argument Joachim is proposing. But his unwillingness to let himself be pinned down on this matter is as characteristic of his method of debate as Moore's argument was of his method. For Joachim says at one point:

Mr. Moore is saying in effect "Give me a plain answer, Yes or No". But I contend, and have tried to establish in my book, that the subject-matter of Metaphysics does not admit this kind of treatment. (Joachim 1907: 411)

Joachim's response here shows not only the gap between himself and Moore, but also that between his kind of idealist philosophy and the more hard-edged approach of his analytic critics. It is hard not to prefer the latter. For although, as we shall see, their own discussions of truth are far from unproblematic, at least their method of argument, which recognises the demand for "a plain answer, Yes or No," enables discussion and debate to move forward.

5.5 MOORE'S "NIGHTMARE"

As we saw above, Bradley's account of the full achievement of truth leads him to a mystical vision of the Absolute. By contrast, according to John Maynard Keynes "Moore had a nightmare once in which he could not distinguish propositions from tables" ("My Early Beliefs"—Keynes 1972: 444). In fact Moore's "nightmare" was a central theme

of his early philosophy. In August 1898 he wrote to his friend Desmond MacCarthy "I have arrived at a perfectly staggering doctrine . . . An existent is nothing but a proposition."²

Moore's route to this doctrine lies through his critical reaction to what he took to be the account of logical ideas that Bradley had advanced in *The Principles of Logic*. Moore first developed this reaction in the dissertation which gained him a Prize Fellowship in 1898 at Trinity College, Cambridge, and he published the core of his new position in his paper "The Nature of Judgment" (*Mind* 1899) (for details of the differences between these texts, see Moore, "Editors' introduction" (2011: esp. lxxv–lxxix). To understand Moore's position, we need to return briefly to Bradley and *The Principles of Logic*. Bradley here treats logical ideas, the ideal contents of thought, as "abstractions," aspects of mental existence which are separated off for the purposes of discursive thought and argument but which are referred, or ascribed, to reality in judgment, so that where a judgment is true, its ideal content is indeed "fact" and the abstract separation of content from existence is overcome. Moore latches on to Bradley's use of the term "abstraction" and takes him to be advancing an empiricist thesis to the effect that logical ideas are formed by a psychological operation of abstraction from experience; and he then criticizes Bradley on the grounds that any such psychological procedure must be guided by an antecedent understanding of that which is to be abstracted from experience, so that this cannot be the way in which the ideas which enter into this understanding and guide the procedure are first formed (Moore 2003: 3). This familiar rationalist argument is effective against the empiricist view attributed by Moore to Bradley, but in a letter to Moore, Bradley explained that Moore had misunderstood him.³ His position was certainly not the empiricist one Moore supposed (indeed Bradley himself had criticized this position at length in *The Principles of Logic*): his view was just that logical ideas are the aspects of our total mental content through which our thoughts have a meaning, a meaning such that where a thought is true its ideal content is a fact. So his talk of "abstraction" was not an account of the formation or creation of ideas, but only of a separation of aspects that were already there.

Bradley's letter indicates that Moore had misunderstood him, though his discussion still leaves many questions concerning the status of "abstract" logical ideas unanswered. However, what is important here is the conclusion that Moore draws from his criticism of Bradley, namely that ideas, or "concepts" as he prefers to call them, are fundamental and not constituted by abstraction or any other process (Moore 2003: 8). In the course of his critical discussion of Bradley Moore alludes to Plato, and one might therefore think that he is espousing a kind of Platonism concerning concepts, comparable perhaps to Frege's conception of senses as a "third realm" ("Thoughts," Frege 1984: 363). However as Moore's discussion continues it emerges that his position is much more radical; to put the point in Fregean terms, Moore has no distinction between sense and reference, so

² This letter is preserved in the Moore archive in Cambridge University Library.

³ Letter likewise preserved in the Moore archive.

that although his concepts are meanings, and thus constituents of propositions, they are also constituents of things such as a rose. Indeed, as he puts it, “the concept turns out to be the only substantive or subject” (Moore 2003: 18). It is, I suggest, best to think of Moore’s concepts as fundamental properties; for one can then understand how it might be that, as he puts it, “a thing becomes intelligible first when it is analysed into its constituent concepts” (Moore 2003: 8)—one has only to think of the familiar “bundle theory” of objects. Equally the role of concepts as meanings can be viewed as a rejection of a sense/reference distinction with respect to these fundamental properties (it is worth noting in this context that although current usage of the term “concept” is typically Kantian, such that concepts are thought of as ways of classifying or thinking of things, for Frege a concept—*Begriff*—is the “reference” (*Bedeutung*) of a predicate, and thus best thought of as a property). But what of propositions, the objects of belief and the like which are constituted by concepts? Moore says that “a proposition is nothing other than a complex concept” (Moore 2003: 5), and the key point that one needs to bear in mind here is the “staggering doctrine” from which we started—“an existent is nothing but a proposition.” So propositions must be understood in such a way that existents just are true existential propositions. Hence one makes best sense of Moore’s position by thinking of his propositions as possible states of affairs, constituted from properties, and recognizing that he does not distinguish between an existent, such as “this paper,” and its existence, the existence of this paper, which is also the true proposition that this paper exists.

Truth has now entered into the specification of this position; but what is Moore’s account of truth? Moore acknowledges that “it is at first sight tempting to say that the truth of a proposition depends on its relation to reality” (Moore 2003: 5). But having presented his accounts of existents as true existential propositions he argues that this temptation must be resisted:

It is similarly impossible that truth should depend on a relation to existents or to an existent, since the proposition by which it is so defined must itself be true, and the truth of this can certainly not be established, without a vicious circle, by exhibiting its dependence on an existent. (Moore 2003: 6)

Given Moore’s account of existents as true existential propositions, his argument here is sound, though to understand it correctly one needs to recognize that when Moore says “the proposition by which it is so defined must itself be true,” his use of the pronoun “it” should be understood as tied to the phrase “an existent” and not to the term “truth” in a way which is also syntactically legitimate. But what then is Moore’s alternative account of truth? He writes,

A proposition is constituted by any number of concepts, together with a special relation between them; and according to the nature of this relation the proposition may be either true or false. What kind of relation makes a proposition true, what false, cannot be further defined, but must be immediately recognised. (Moore 2003: 5)

Moore also says, concerning the proposition “This paper exists”:

But if it is true, it means only that the concepts, which are combined in specific relations in the concept of this paper, are also combined in a specific manner with the concept of existence. That specific manner is something immediately known, like red or two. It is highly important, because we set such value upon it; but it is itself a concept. (Moore 2003: 6)

So truth and falsehood are “special” relations which combine the concepts inherent in a proposition into a proposition. These relations can be immediately recognized, but cannot be defined—though no reason is given why falsehood should not be defined in terms of truth, as what is not true. Truth however is the concept which we value, and it is absolutely fundamental; in particular “truth cannot be defined by a reference to existence, but existence only by a reference to truth” (Moore 2003: 6).

Although Moore’s pluralist atomism is manifestly very different from Bradley’s holistic monism, it is worth noting that there are some significant similarities between them. An initial point concerns the denial of contingency: just as Bradley’s holism led him to think that all truths are connected in such a way that they necessitate each other, Moore held that all judgments are “equally necessary whether [they] be true or false” (Moore 2003: 18). In Moore’s case this necessity does not arise from the holistic interconnections Bradley postulates; instead Moore’s claim, I think, rests on his conception of truth and falsehood as the relations which connect the concepts in a proposition, so that where concepts are combined in the “true” way in a proposition such as “This paper exists,” the very same proposition would not have been constituted had the concepts been combined in the “false” way. In fact, however, this conception of truth and falsehood as the relations which combine concepts in a proposition is problematic, since, on the face of it, it is one thing for there to be a proposition such as “Green was older than Bradley,” and quite another for it to be true or false. So although Moore does not, I think, explicitly modify his conception of truth and falsehood in such a way that they are just simple properties of propositions whose constituents are combined without reference to these properties, this change, which, as we shall see below, Russell soon introduced, would be sensible, and would also remove the grounds for the unattractive thesis that all judgments are “equally necessary whether [they] be true or false.”

Moving on now to truth itself, Moore, like Bradley, emphatically rejects any conception of truth as involving correspondence between a representation, mental or linguistic, and the world. In the brief entry on “Truth and Falsehood” which Moore wrote for Baldwin’s *Dictionary of Philosophy and Psychology* (Moore 1902) he is particularly clear on this point: “It is the impossibility of finding any such difference between a truth and the reality to which is supposed to correspond which refutes the theory” (Moore 1993: 21). Thus, Moore continues, “the truth that I exist differs in no respect from the corresponding reality—my existence” (ibid.). So Moore’s rejection of a correspondence theory rests upon his affirmation of the identity between truth and reality—a thesis we have also found in Bradley. Having noted this striking similarity between their accounts

of truth, however, two important differences should be acknowledged: first, whereas for Bradley the ultimate identity between truth and reality provides a definition of absolute truth, for Moore things are the other way around, as he himself observes:

So far, indeed, from truth being defined by reference to reality, reality can only be defined by reference to truth: for truth denotes exactly that property of the complex formed by two entities and their relation, in virtue of which, if the entity predicated be existence, we call the complex real—the property, namely expressed by saying that the relation in question does truly or really hold between the entities. (Moore 1993: 21)

Second, for Moore this identity between truth and reality primarily concerns simple existential propositions, whereas for Bradley, of course, it just concerns reality as a whole—the Absolute.

The last point takes us back to the point of radical disagreement between Moore and Bradley—Moore's pluralist atomism versus Bradley's holistic monism. One implication of this disagreement is that whereas Bradley allows that there are degrees of truth, Moore rejects any such conception—"There are, properly speaking, no degrees of truth and falsehood" (Moore 1993: 20). But, at least in his early writings, Moore does not attempt to explore the basis of this disagreement. In the discussion above of Bradley I argued that Bradley's position is bound up with his treatment of relations, and that although Bradley's assumptions concerning this topic can be challenged, his arguments on this topic merit serious attention. But Moore does not attempt to deal at all with Bradley's arguments. For example, Moore's account of truth as a special relation which combines the concepts inherent in a proposition into a proposition is precisely open to Bradley's challenge as to how this combining is achieved—whether we think of the proposition as a possible state of affairs or as a judgment. Moore, in effect, just assumes that Bradley's challenge can be met, and sets out his pluralist alternative without any attention to the reasons Bradley has given for holding that such an alternative is untenable. For example, according to Moore,

Concepts are possible objects of thought; but that is no definition of them. It merely states that they may come into relation with a thinker; and in order that they *may* do anything, they must already *be* something. It is indifferent to their nature whether anybody thinks them or not. (Moore 1993: 4; emphasis in original)

For Bradley, this account of thought as an external relation between a thinker and what she thinks would be altogether untenable: the relation between a thinker and her thought needs to be grounded within a non-relational system that encompasses both. No doubt Bradley's claim here is extravagant and unwarranted, but it is disappointing that Moore makes no attempt in his early writings to explain why he rejects it.

Many years later Moore did make such an attempt, in his paper "External and Internal Relations" (1919). Much of this paper is given over to a subtle account of the distinction

between external and internal relations, which leads Moore to articulate distinctions that are important in the development of modal logic. On the topic of internal relations Moore notes that there is an important distinction between the implication of “Leibniz’s Law” that

(A) aRb entails that (if Not (cRb) then ($a \neq c$))

and the claim which is characteristic of the treatment of “ R ” as an internal relation, that

(B) aRb entails that (Not (cRb) entails ($a \neq c$))

The first of these is, Moore thinks, obviously true; but the second, which, Moore thinks, is liable to be confused with it by those who accept the axiom of internal relations, does not follow from it and needs to be distinguished from it. The significance of this distinction is that although (A) implies

aRb entails that (if $a = a$ then aRb)

one cannot infer that aRb entails that it is a necessary truth that aRb despite the fact that it is a necessary truth that $a = a$. By contrast, (B) implies

aRb entails that ($a = a$ entails aRb)

and given this proposition and the necessary truth of $a = a$, one can now infer that aRb entails that it is a necessary truth that aRb . So, for example, it follows from (B) that wherever two things, a and b are such that a is located to the left of b it could not have been the case that a was not on the left of b .⁴ On the face of it, this is absurd and Moore seems to think that implications of this kind are a *reductio* of the axiom of internal relations (Moore 1993: 97–8).

In fact the dialectical situation here is more complex than Moore acknowledges, for, despite his unease about Russell’s attribution to him of the “axiom of internal relations,” Bradley does make precisely the claim about the necessity of spatial relations identified above. He attempts to make it plausible by means of a distinction between things as they really are, which include their spatial relations, and the abstracted aspects of things which we often mistake for them, which do not include spatial relations and give rise to the appearance of contingency in this respect (Bradley 1897: 575–8). It is, however, hard to make much sense of this, and in this respect, therefore, it is fair to conclude that in this later paper Moore does engage effectively with one aspect of Bradley’s treatment of relations, though he does not deal with the broader issue of the unity of relations and Bradley’s regress argument.

⁴ Moore’s example is more complex than that given here, though not different in principle.

5.6 RUSSELL—AND BRADLEY’S REVENGE

Russell always acknowledged the impact of Moore’s “The Nature of Judgment” on his thought, and in the preface to *The Principles of Mathematics* (1903) he wrote that “On fundamental questions of philosophy, my position, in all its chief features, is derived from Mr G. E. Moore” (Russell 1903: xviii). There were in fact some disagreements between them, for example concerning particulars and universals; furthermore in their critical reaction to Bradley, it was Russell who took the lead by arguing explicitly against the “axiom of internal relations.” On truth, however, Russell did at first follow Moore, taking it that

It may be said—and this is, I believe, the correct view—that there is no problem at all in truth and falsehood; that some propositions are true and some false, just as some roses are red and some white . . . But this theory *seems* to leave our preference for truth a mere unaccountable prejudice, and in no way to answer to the feeling of truth and falsehood. (“Meinong’s Theory of Complexes and Assumptions” (1904), reprinted in Russell 1994: 473; emphasis in original)

However by 1907 Russell had come to think that this simple theory really did not deal properly with “the feeling of falsehood.” For on this theory, as he put it, “there are not only mistaken beliefs, but also non-facts, which are the objectively false objects of mistaken beliefs” (Russell 1907: 45). These “non-facts,” or “fictions,” as he also calls them, are false propositions, and the trouble is that, on this theory, they have exactly the same ontological status as true propositions, that is, facts. But, as he put it a couple of years later, “it is very hard to believe that there are such objects as ‘that Charles I died in his bed’” (Russell 1910: 151) given that there was no such event as Charles I’s death in his bed.

Russell’s reaction to this situation was to propose an alternative strategy, according to which, say, Russell’s belief that Green was older than Bradley is not construed as belief in a proposition, the proposition that Green was older than Bradley, but is instead taken to involve a “multiple relation” between Russell, the subject of the belief, and the object terms of the belief, namely Green, Bradley and the relation of being older than (Russell 1910: 155). Because this strategy deconstructs the proposition as an object of belief, it removes the old problem concerning the status of false propositions. But it now gives rise to a new problem concerning the truth of the belief, namely how it is to be determined that the truth of the belief consists in Green’s being older than Bradley, as opposed to Bradley’s being older than Green. That is, how is it to be determined which terms of the multiple relation, belief, and in which order, specify the conditions for the truth of the belief.

Although he does not acknowledge this, it should have been obvious to Russell that this new problem is just a variant form of Bradley’s old challenge, to the effect that the unity of a relationship, or a relational judgment, cannot be secured simply by specifying a relation and its terms. Russell, unlike Moore, had been sensitive from the start to

this challenge. Early in 1899, soon after he had read Moore's "The Nature of Judgement," Russell gave a paper to the Cambridge Moral Sciences Club on "The Classification of Relations," in which he defends the irreducible reality of relations but ends by restating Bradley's challenge:

When two terms have a relation, is the relation related to each? To answer affirmatively would lead at once to an endless regress; to answer negatively leaves it inexplicable how the relation can in any way belong to the terms. I am entirely unable to solve this difficulty. (Russell 1990: 146)

In *The Principles of Mathematics* Russell does present himself as having a solution to the difficulty, namely that the regress is harmless (1903: 99–100). His view seems to be that he can solve Bradley's challenge by distinguishing between "a relation in itself" and "a relation actually relating" (1903: 49), and rely on the latter to provide an affirmative answer to the question above because "a relation actually relating" sustains the unity of a relational proposition without any need for further relations, so that even if further relating relations are introduced to spell out the connections here, the regress is harmless. But he acknowledges that he has no clear account of the key distinction between a relation in itself and a relation actually relating; he suggests that "a relating relation is distinguished from a relation in itself by the indefinable element of assertion which distinguishes a proposition from a concept" (1903: 100) but he is not confident about this, and since it would leave him with no account of the unity of unasserted propositions or other complexes such as Green's being older than Bradley it does not appear likely to be satisfactory. So this early dismissal of Bradley's challenge is unpersuasive.

When Russell introduces his multiple-relation theory, he recognizes the challenge posed by beliefs which concern relationships that are not symmetrical: the belief that A loves B is clearly distinct from the belief that B loves A, so it is not enough to say that the truth of the belief requires that the relation involved as an object term, love, actually relate the other terms. It has to relate them in the right way. Russell attempts to deal with this by saying that within the belief, the relation involved is given as having a specific "sense," for example as going from A to B, and thus that the truth of the belief requires that the relation involved should relate the other object terms with the same sense—such that A loves B (Russell 1910: 158). This sounds fine: but now Russell has given himself the task of explaining how to square the essential role of the "sense" of a belief with the conception of belief as a multiple relation; and this is where Bradley gets his revenge. For the sense which connects terms in the right way to a relation which is itself an object term of the multiple-relation belief looks like the extra links which are postulated to solve the unity of a relational judgment but which just head one off on an endless regress. Russell attempted to deal with this issue in his 1913 "Theory of Knowledge" (Russell 1984), but fortunately it is not necessary to discuss here this ill-fated endeavor whose increasingly complex formulae remind one of the epicycles of Ptolemaic astronomy as Russell struggles to provide an account of the sense of a relation (see e.g. the discussion in Russell 1984: 147). For Russell himself abandoned the theory in the light

of Wittgenstein's criticisms of it. It is not clear what Wittgenstein's criticisms were, and they do not appear to focus on Russell's difficulties in accommodating the sense of a relation within the terms of a multiple relation such as belief. Nonetheless it is notable that when Russell returned to this topic in his 1917/18 lectures on "The Philosophy of Logical Atomism" he wrote,

There are really two main things that one wants to notice in this matter that I am treating of just now. The first is the impossibility of treating the proposition believed as an independent entity, entering as a unity into the occurrence of the belief, and the other is the impossibility of putting the subordinate verb on a level with its terms as an object term in the belief. That is a point in which I think the theory of judgment which I set forth once in print some years ago was a little unduly simple, because I did then treat the object verb as if one could put it as just an object like the terms, as if one could put "loves" on a level with Desdemona and Cassio as a term for the relation "believe". (Russell 1986: 199)

Russell had called the account of truth associated with his multiple-relation theory a "correspondence" theory (Russell 1910: 158), on the grounds that where a belief is true its object terms, including a relation somehow connected to its terms by a definite sense, "correspond" to a complex in which the very same terms occur, but now with the relation actually relating them in the order specified by the sense. This is far removed from the kind of correspondence theory criticised by Bradley and Moore, since the object terms of the belief do not represent, or "copy," objects and relations—they are the very objects and relations involved; so the resulting account of truth (assuming the difficult problem of sense could be solved without requiring a radical revision of the theory) would in fact be a close cousin of the identity theory of truth. But having abandoned the multiple-relation theory at least partly because of the problem of sense, Russell switched in his lectures on "The Philosophy of Logical Atomism" to a quite different position: that the fundamental bearers of truth and falsehood, which he still calls "propositions," are complex "symbols" (166), typically statements. As Russell acknowledges, this was a very radical shift in his philosophy.

It is not necessary, or appropriate to discuss this change in any detail here; but in order to bring this discussion of truth to a conclusion it is worth briefly looking at Russell's new account of truth. In his lectures he declares

That the components of the fact which makes a proposition true or false, as the case may be, are the *meanings* of the symbols which we must understand in order to understand the proposition. (Russell 1986: 175; emphasis in original)

Russell goes on to qualify this last statement: it applies primarily to atomic propositions, and the truth of molecular propositions is a different matter. But with respect to atomic propositions Russell is happy to talk of the "corresponding" facts (Russell 1986: 175): where there is a correspondence between the constituent symbols of a proposition and the components of a fact which are the meanings of the symbols, then that

fact makes the proposition true. This is now a genuine correspondence theory of truth, though it retains a residue of the old identity theory of truth and reality in the conception of a logically perfect language:

I shall try to persuade you that in a logically correct symbolism there will always be a certain fundamental identity of structure between a fact and the symbol for it; and that the complexity of the symbol corresponds very closely with the complexity of the facts symbolized by it. (Russell 1986: 175)

This “fundamental identity of structure between a fact and the symbol for it” shows Russell’s loyalty to the old Moorean identity theory of truth despite his shift to symbolic propositions. But the benefits of the new position are immediate: it enables him to accommodate false propositions without embarrassment—there is no need to think of them as corresponding to non-facts. Equally the old issues concerning relations which plagued the multiple-relation theory can be set to one side, to be handled within an ontological theory concerning the “components of a fact” and a theory of meaning concerning the corresponding “components of a proposition.” There have been, and still are, plenty of different approaches to these issues; but they no longer dominate discussions of truth.

So this position which Russell arrives at over the course of his twenty-year debate with Bradley, Joachim, Moore, and others on this topic since 1897 is the place to bring an end to this discussion of the tangled debates between the British idealists and their analytic critics. Russell’s “symbolic” (linguistic) turn does decisively move the argument forward by freeing the discussion of truth both from the speculative metaphysical complexities of the idealists and from the unsatisfactory positions he and Moore had at first adopted in the course of their liberation from idealist metaphysics.

REFERENCES

- Boole, G. (1854). *The Laws of Thought*. London: Walton and Maberly.
- Bosanquet, B. (1895). *The Essentials of Logic*. London: Macmillan.
- Bradley, F. H. (1897). *Appearance and Reality*, 2nd edn. London: George Allen & Unwin.
- Bradley, F. H. (1914). *Essays on Truth and Reality*. Oxford: Clarendon Press.
- Bradley, F. H. (1922). *The Principles of Logic*, 2nd edn. London: Oxford University Press.
- Bradley, F. H. (1935). *Collected Essays*. Oxford: Clarendon Press.
- Frege, G. (1984). *Collected Papers on Mathematics, Logic, and Philosophy*, ed. B. F. McGuinness. Oxford: Blackwell.
- Frege, G. (1953). *The Foundations of Arithmetic*, trans. J. L. Austin. Oxford: Blackwell.
- Green, T. H. (1883). *Prolegomena to Ethics*, ed. A. C. Bradley. Oxford: Clarendon Press.
- Green, T. H. (1886). *Works of T. H. Green*, ed. R. L. Nettleship, vol. 2. London: Longmans, Green.
- Ishiguro, H. (1972). “Leibniz’s Theory of the Ideality of Relations,” in H. Frankfurt (ed.), *Leibniz: A Collection of Critical Essays*. New York: Doubleday, 191–214.
- Joachim, H. (1906). *The Nature of Truth*. Oxford: Clarendon Press.
- Joachim, H. (1907). “A Reply to Mr. Moore,” *Mind* 16: 410–15.

- Keynes, J. M. (1972). *Essay in Biography*, Collected Writings vol. 10, extended edn. Cambridge: Cambridge University Press.
- Keynes, J. N. (1884). *Formal Logic*. London: Macmillan.
- Locke, J. (1975) [1689]. *An Essay Concerning Human Understanding*, ed. P. Nidditch. Oxford: Oxford University Press.
- Moore, G. E. (1899). "The Nature of Judgment," *Mind* 8: 176–93. Reprinted in Moore 1993: 1–19.
- Moore, G. E. (1902). "Truth and Falsehood," in J. Baldwin (ed.), *Dictionary of Philosophy and Psychology*, vol. 2. London: Macmillan, 716–18. Reprinted in Moore 1993: 20–2.
- Moore, G. E. (1907). "Mr. Joachim's *Nature of Truth*," *Mind* 16: 229–35.
- Moore, G. E. (1919). "External and Internal Relations," *Proceedings of the Aristotelian Society* 20: 40–62. Reprinted in Moore 1993: 79–105.
- Moore, G. E. (1993). *Selected Writings*, ed. T. Baldwin. London: Routledge.
- Moore, G. E. (2011). *Early Philosophical Writings*, ed. T. Baldwin and C. Preti. Cambridge: Cambridge University Press.
- Russell, B. (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- Russell, B. (1906). "The Nature of Truth," *Mind* 15: 528–33.
- Russell, B. (1907). "On the Nature of Truth," *Proceedings of the Aristotelian Society* 7: 28–49.
- Russell, B. (1910). *Philosophical Essays*. London: George Allen & Unwin.
- Russell, B. (1937) [1900]. *A Critical Exposition of the Philosophy of Leibniz*, 2nd edn. London: George Allen & Unwin.
- Russell, B. (1984). *The Collected Papers of Bertrand Russell*, vol. 7. London: Routledge.
- Russell, B. (1986). *The Collected Papers of Bertrand Russell*, vol. 8. London: Routledge.
- Russell, B. (1990). *The Collected Papers of Bertrand Russell*, vol. 2. London: Routledge.
- Russell, B. (1994). *The Collected Papers of Bertrand Russell*, vol. 4. London: Routledge.
- Russell, B. (1995). *My Philosophical Development*, revised edn. London: George Allen & Unwin.

CHAPTER 6

JUDGMENTS, FACTS, AND PROPOSITIONS

*Theories of Truth in Russell, Wittgenstein,
and Ramsey*

PETER SULLIVAN AND COLIN JOHNSTON

6.1 INTRODUCTION

OUR aim in this chapter¹ is to outline a story that ought to be familiar and unsurprising, one that traces the fate of the correspondence theory of truth from its adoption by Russell in “On the Nature of Truth and Falsehood” (1910) to its repudiation by Ramsey in “Facts and Propositions” (1927). Central episodes in this story are indeed very familiar. But commonly held views of them, when placed one after the other, make for a story that is more surprising and less coherent than it should be: slightly misplaced emphasis at the beginning, regarding Russell’s reasons for adopting his new theories of judgment and truth, sets things off in a direction that leads to simple error in the middle, regarding Wittgenstein’s views in the *Tractatus*; this error then calls for a sudden and inexplicable plot-twist in the transition to the final chapter, regarding Ramsey’s position, which in consequence is bungled.

At the turn of the twentieth century, and through the period of *The Principles of Mathematics* (1937 [1903]) and “On Denoting” (1905a), Russell endorsed a theory of judgment

¹ While the chapter is jointly written, so that we are severally and collectively responsible for any mistakes in it, it may be helpful to note that most of the central section on Wittgenstein (sections 6.3.1–6.3.3) was first drafted by Johnston and revised by Sullivan, while the reverse holds of other sections. In sections 6.3.1–6.3.3, use is made of material from Johnston’s “The Picture Theory,” forthcoming in H.-J. Glock and J. Hyman (eds.), *The Blackwell Companion to Wittgenstein*; some paragraphs in sections 6.4.2–6.4.3 overlap with Sullivan’s “An Introduction to ‘Facts and Propositions,’” published online in an Aristotelian Society “virtual symposium” in April 2013; we are grateful to the editors and publishers for permission to make use of this material. We should also like to thank Alan Millar and Michael Potter for comments on a partial draft.

or belief as consisting in a dyadic relation between a judging mind and a unitary objective entity embodying the content of the judgment, which he termed a “proposition.” The truth or falsity of a judgment is on this view derivative from the truth or falsity of the proposition judged. Further, truth and falsity are held to be simple, unanalyzable properties of propositions. In writings from 1904 onward Russell presents various reasons for dissatisfaction with this early theory, and by 1910 they have led him to abandon it. Some of these reasons are acknowledged to be suasive considerations only; mostly these are elaborations of the intuitive appeal of the notion that “true propositions express *fact*, while false ones do not,” and that “when a proposition is false, *something* does not subsist which would subsist if the proposition were true” (1904: 473; emphasis in original). Others focus on difficulties internal to the early theory, which are said to be “grave” (1904: 46), “more fatal” (1910: 119), and ultimately to render that theory “impossible” (1912: 193). Our contention in section 6.2 is that an explanation of Russell’s change of mind should emphasize considerations of the second kind over the first. Treatments which emphasize the first will tend to present the desire to accommodate a “correspondence” intuition as central to Russell’s motivation. They will in consequence tend to obscure the fact that the association between the two components of Russell’s new view, the “multiple-relation theory” of judgment and the “correspondence theory” of truth, is optional. By this we mean that the association is optional *in itself*. It might have been inevitable *for Russell*, but if so then the reason lies in more basic commitments that persist through the change in his theories of judgment and truth, and merely assume a new form in his post-1910 theory of complexes.

In 1913 Wittgenstein famously criticized Russell’s new theory of judgment; and it has been compellingly argued (Ricketts 1996; Potter 2009) that this criticism contains the seeds of the “picture theory” Wittgenstein would present in the *Tractatus*. It is certainly true that the picture theory is offered as a corrective to Russell’s views. However, neglect of the role of Russell’s basic commitments about complexes, both in determining the specific form of his multiple-relation theory of judgment and in associating with it a correspondence theory of truth, leads to misunderstandings of what Wittgenstein retained from Russell’s theories and what he rejected. Probably the most widely held views are that Wittgenstein repudiated the multiple-relation theory, but that he retained its associated correspondence theory of truth. In section 6.3 we argue that the opposite is the case. We hold, first, that the picture theory is itself a multiple-relation theory, though one implemented without the framing commitments of Russell’s view of complexes; and secondly, that by freeing itself from those commitments about complexes Wittgenstein’s theory of judgment has no need for, and indeed allows no room for, a correspondence theory of truth.

Ramsey’s “Facts and Propositions” (1927) is the third of the three great published articles in which Ramsey expounds and builds on Wittgenstein’s work. It addresses important gaps in the picture theory, which Ramsey had identified as early as in his Critical Notice of the *Tractatus* (1923: 275–7). Famously, Ramsey denies in this essay the need for any such substantive theory of truth as the correspondence theory, concluding early on that “there is really no separate problem of truth” (1927: 142). Less famously, Ramsey

sets the theoretical context for his innovations by beginning the article with a compelling argument in favor of a multiple-relation theory. According to the widely held views just mentioned, these contentions would represent fundamental differences between Ramsey and Wittgenstein. Holders of such views must therefore resort to desperate contortions to explain, or else must simply ignore, Ramsey's closing acknowledgment to Wittgenstein: "Everything that I have said is due to him," Ramsey says, excepting from this only the "pragmatist" elements of his theory that he judged necessary to "to fill up a gap in his [Wittgenstein's] system" (1927: 155). In section 6.4 we present a sketch of Ramsey's views of judgment and truth that straightforwardly accords with this judgment.

We remarked at the outset that the story we aim to outline here ought to be familiar and unsurprising. One reason for this is that the story was already well told by Ramsey himself. Hence Ramsey will be a point of reference in earlier sections, as well as in the section devoted specifically to his own views.

6.2 RUSSELL

6.2.1 The early theory

Russell's early theory of judgment and truth can be regarded as the upshot of two commitments. The first is an expression of the form Russell's realism took at this period, and holds that cognition is an external relation between a mind and an objective reality. Knowledge, Russell insists, is of nothing less than a mind-independent fact. In being known, the known fact of course enters into a relation with a mind, but the existence of the fact is in no way dependent on that relation, nor is its nature in any way altered by it (1904: 461–3; 1905b: 501; 1906–7: 45). The second commitment, which extends through Russell's change of mind, is that a false belief is still a belief: that someone believes that so-and-so is a state of affairs of a certain kind that is exemplified equally by true and false beliefs (1904: 467; 1910: 120). Again, Russell often presents this as an aspect of realism, though this time of a more common-sensical kind. If John believes that his train will be late arriving, then whether his belief is true or false depends on how things stand somewhere up the line, and is not to be discerned by an internal scrutiny or analysis of his state of believing. Truth and falsity are thus "extrinsic," not "internal" properties of belief (1912: 189–90, 201). It follows that true and false beliefs must be analyzed in the same way, and are ontologically on a par. Since the object of a true belief is a mind-independent fact, a false belief must similarly have as its object a similarly mind-independent entity, something that is just like a fact with the sole exception that it is not true. This something is a "proposition."

The above exposition follows Russell's typical practice in seeming initially to accord some priority to the notion of a fact, the object of a true belief (e.g. 1904: 467);

propositions enter the account only secondarily, as the objects of belief considered independently of whether they are true or false. The upshot, however, is that facts have no fundamental status in Russell's theory. The parallel in structure and analysis between true and false belief demands that the theory's primary ontological notion is that of a proposition. Then, just as people who happen to have taken degrees are called "graduates," so propositions "are called *facts* when they happen to be true" (1905b: 492; emphasis in original). When they happen instead to be false, they might be called "fictions" (1906–7: 48), though Russell more often and more neutrally calls them "objective falsehoods." The key point, though, is that we need call them neither of these things. To name either a fact or a fiction for what it fundamentally is, without building in extraneous presumptions about it, we should call it a proposition.

This core theory provides for no connected entity which exists when a belief is true but which fails to exist when it is false. On the one hand, because they are abstract entities neither true nor false propositions can strictly be said to "exist," in the narrow sense Russell reserves for that term (that of existing in time); on the other, both true and false propositions equally "subsist" or "have being." Nonetheless, Russell readily acknowledges the appeal of this notion, and admits that it might be accommodated. That is, we might recognize "an allied concept of an entity, got grammatically [from a proposition] by turning the verb into a verbal noun, . . . [which] denotes an entity when the proposition is true, but not when it is false" (1905b: 506). But this could only be a supplement to the core theory. If we do recognize such an entity, designated (for instance) by "the death of Caesar," we must still insist that this entity "is distinct from the proposition" that Caesar died. It does not enter into the analysis of judgment, and is not what is true or false (1904: 471).

The core theory, then, is that belief or judgment consists in a subject's entering into a dyadic relation to an objective proposition. A proposition is an abstract, complex entity. Its constituents are, like the proposition itself, entirely mind-independent. They are the actual entities with which the proposition is concerned; the proposition that Andrew loves Beatrice, for instance, has as its constituents Andrew, Beatrice, and the relation of loving (1905b: 494). Since the identity of a proposition is not determined solely by that of its constituents, but depends also on their configuration, a proposition is not a mere collection or "aggregate" (1904: 437). It is a unity of a special kind, a kind that Russell finds it hard to characterize. It consists, he suggests, not in the presence of a relational constituent, but "simply and solely in the fact of relatedness in a certain way" (1904: 437); or, as he elsewhere puts it, "a proposition consists of a relation between its terms" (1905b: 495), not merely of certain terms and a relation.

Truth and falsehood Russell then holds to be simple, unanalyzable properties of propositions, from which the truth or falsity of beliefs or judgments derives. This view of the primitiveness of truth certainly coheres with, and is typically presented by Russell in close association with, the theory of judgment just outlined (1904: 473–4; 1906–7: 48). However it is not, either in Russell's view or in fact, a direct consequence of that theory. Russell offers (1905b: 493–4) an independent, general argument—strikingly reminiscent of Frege's (1979: 128–9) though apparently deriving from Moore's (1899: 64–5)—for the

indefinability of truth; and he argues in detail against coherence and pragmatist analyses of the notion (1905b; 1906–7). What does follow from the considerations already presented is that truth is not to be explained by the notion of a fact, which is instead dependent on it.

6.2.2 Problems for the early theory

Even in the essays that advocate it Russell recognizes that the early theory faces serious problems, and as early as 1904 he formulates an alternative to it. According to this alternative, which represents the embryonic form of his multiple-relation theory, belief would be conceived, not as a state with a single complex object, but instead as a complex state with a multiplicity of objects (1904: 468). Although Russell develops what he then takes to be conclusive objections to it (1904: 468–9),² this alternative reappears almost unchanged in “The Nature of Truth” (1906–7) as the view that belief is “not a single idea” but “a complex of ideas” (1906–7: 47), or that “a belief . . . will not consist of one idea with a complex object, but will consist of several related ideas” (1906–7: 46). In favor of this alternative Russell observes, first, that it removes a commitment to “objective non-facts”; for, as we saw, if true beliefs have unitary objects, so must false beliefs, and “if beliefs always have objects, it follows that there are objective non-facts” (ibid.). His second observation is that the alternative would resolve paradoxes, such as the *liar*, that are generated by quantification over propositions—by the clean but drastic device of eliminating from his ontology altogether the propositions quantified over (1906–7: 49). Remnants of the 1904 objections persist, however, and so at this stage Russell declines to choose between his early theory and this alternative (ibid.).

By 1910, when Russell came to republish this paper in the collection *Philosophical Essays*, the issue had been decided in favor of the alternative theory. Hence only the first two, critical parts of the essay were reprinted, under the title “The Monistic Theory of Truth,” while the inconclusive Part III was replaced by a new essay, “On the Nature of Truth and Falsehood” (1910). Most of this section will be concerned with the reasons Russell gave there for this decision. First, though, we need a brief statement of the new theory.

The theory of judgement that I am advocating is, that a judgement is not a dual relation of the mind to a single Objective, but a multiple relation of the mind to the various other terms with which the judgement is concerned. Thus if I judge that A loves B, that is not a relation of me to “A’s love for B”, but a relation between me and A and love and B. (1910: 122)

² The decisive premise in Russell’s objection is that “the presentation of a relation is not a relation.” This serves to undermine the attractive suggestion that “the presentation of ‘*a* and *b* related by *R*’” should consist in “the presentations of *a* and *b* related by the presentation of *R*” (1904: 468). Wittgenstein and Ramsey were later to reject the premise of Russell’s argument, holding that the presentation of a relation *is* itself a relation, and so they could simply accept the attractive suggestion.

In the early theory, in judging that *a* loves *b* a subject stands in relation to a single complex entity, a proposition, of which *a*, love, and *b* are constituents. Schematically, the fact of *s*'s so judging has the form: $J(s, \langle aLb \rangle)$. The new theory recognizes no such entity as $\langle aRb \rangle$: "the mere fact that the judgement occurs does not involve any relation between its objects *A* and love and *B*" (1910: 122). Schematically, its form is: $J(s, a, L, b)$.

Building on this account of the form of a judgment Russell offers a correspondence theory of what it is for the judgment to be true. Given any judgment, it should be possible to define the notion of a corresponding complex, whose constituents will be all of the constituents of the judgment excepting the judging subject and the judgment relation. Continuing with the previous example of *s*'s judgment that *a* loves *b*, the constituents of this corresponding complex will be *a*, the relation of loving, and *b*; and the corresponding complex will be one in which *a* is related to *b* by the relation of loving.³ The judgment will be true if there exists such a corresponding complex, and false if the corresponding complex does not exist.

In summary, according to Russell's early theory,

- (i) for each fact of judgment there is a corresponding complex entity, the judged proposition;
- (ii) truth and falsity are primarily features of propositions, and only derivatively of judgments which have those propositions as their objects.

On the new theory,

- (i) truth and falsity are features only of judgments;
- (ii) only in the case of a true judgment does there exist a corresponding complex entity, the fact that is judged to obtain.

Thus the general category of propositions is eliminated, and the only complex entities recognized are facts.

It is also interesting to compare this new theory with what we might call the proto-multiple-relation theory of the 1904 and 1906–7 papers. The suggestion then was that a judgment is "a relation of presentations" (1904: 468), or "a complex of ideas" (1906–7: 47). In the 1910 statement of the theory there is no mention of "presentations" or "ideas":⁴ the multiple relation of judgment is said to hold between the judging subject

³ Like Russell's own initial statement of the theory, this skates over complications to do with the ordering of the entities in the corresponding complex.

⁴ At 120–1 Russell does discuss the point that for *s* to be multiply-related by judgment to *a*, *L*, and *b*, *s* must be "conscious of" (or acquainted with) each of the objects *a*, *L*, and *b*. He observes that relations of acquaintance with these objects "separately and severally" will not add up to the judgment, which must be "one single unity." This familiar (Kantian) point was already made by Meinong, and endorsed by Russell earlier in the 1904 essay (439, 455). It does *not* run counter to the proto-multiple-relation theory, which advances the quite different suggestion that judgment is constituted by a *relation between* presentations, or a *complex* of presentations—not by a mere sum of presentations.

and entities the judgment is about. One might suppose that this change in formulation shows Russell's realism reasserting itself after a temporary lapse, with an insistence that the objects of judgment are mind-independent things. But this superficial suggestion is mistaken. The proto-multiple-relation theory is in fact entirely consistent with the 1910 statement, since to be the subject of a complex of presentations of a , L , and b just is one way of being multiply-related to the objects of those presentations, i.e. one way of exemplifying the schema $J(s, a, L, b)$.⁵ The real difference is that the 1910 statement provides *only* this schema, whereas the proto-multiple-relation theory had offered at least some indication of how it might be exemplified. Or, to put things the other way around, the 1910 statement highlights the form of the judgment relation, but it says (even) less than do Russell's preliminary formulations as to what the multiple relation of judgment actually is.

What, then, persuaded Russell to decide in favor of this new theory? His reasons are presented in three dense paragraphs of "On the Nature of Truth and Falsehood" criticizing the view that true and false judgments equally have "objectives." The first considers the notion that the objective of a judgment, for instance, the judgment that Charles I died on the scaffold, might be "the actual event" of Charles I's death on the scaffold. Its relation to the early theory is initially at least unclear. We will return to it after considering Russell's case against the theory that the objective is instead "that Charles I died on the scaffold."

Russell's first objection to this theory is simply that "it is difficult to believe that there are such objects." "It seems evident," he says,

that the phrase "that so-and-so" has no complete meaning of itself, that would allow it to denote a definite object . . . We feel that the phrase "that so-and-so" is essentially incomplete, and only acquires significance when words are added so as to express a judgement, e.g. "I believe that so-and-so" . . . (1910: 119)

Russell is here giving expression to his new doctrine that propositions—or, perhaps better, the phrases previously taken to express propositions—are "incomplete symbols" (Whitehead and Russell 1910: 44). According to this doctrine there are no such entities as the early theory had taken propositions to be. Thus any apparent reference to such an entity must be analyzed away, by giving a different account of the structure of the fact expressed by a statement in which the apparent reference is made. Belief-ascriptions provide one example of a class of such statements, but importantly they are only one example. The demand for analytical reconstrual applies wherever a given fact appears to involve a proposition as a "subordinate complex" (1906–7: 48)—or, as we might now put it, it applies to any sentence apparently involving a sub-sentence or clause that can be supposed false without rendering the whole sentence meaningless. (Obvious further

⁵ This depends on Russell's view that presentation, or its converse acquaintance, is itself a direct cognitive relation between a subject and an independent object.

examples are truth-functionally complex statements, statements of implication, causal statements, statements of probability, and so on.) As Russell had reasonably said when first outlining this approach, whether its general analytical demand can be carried through will turn on many difficult “considerations of detail” (1906–7: 49). It is hard to see how the outcome could be “evident,” or how the “feeling” he mentions might be relevant to it.

The following paragraph, which Russell describes as developing a “second” and “more fatal” objection, in fact presents a range of considerations. They are united in focusing particularly on the early theory’s commitment to *false* objectives.

First, that there should be, in the mind-independent world, entities describable as “objective falsehoods” is, Russell now declares, “in itself incredible.” In arguing for the early theory he had said, oppositely, “it is very hard to believe that nothing is objectively false” (1904: 467). Plainly, someone’s sense of reality can change. So we have here, not so much an argument, as an invitation to diagnose a shift in Russell’s understanding of what an objective falsehood would be. A parallel passage in the later lectures on “The Philosophy of Logical Atomism” is helpful in this. What Russell then “cannot believe,” and indeed finds it “monstrous” to suppose, is that objective falsehoods should “go about the real world,” or that they should exist “in the actual world of nature” as facts do (1918: 196–7). The early theory’s propositions were, we should recall, *abstract* entities. In Russell’s rejection of them they have become concrete entities mysteriously lacking in substance—“curious shadowy things,” the ghosts of departed facts.

Russell now recalls the early theory’s commitment to regard the distinction of truth and falsity as “ultimate and not further explicable.” It cannot accommodate our feeling that, when we judge truly, there exists “some entity ‘corresponding’ in some way to our judgement,” while when we judge falsely there is not, and that the difference between truth and falsity is to be explained by the presence or absence of this “corresponding” entity. With only minor qualifications, the points Russell highlights in this passage are indeed part of the early theory. He now holds that they show the theory to be “unsatisfactory” (1910: 119). We need not judge whether he is right, for it surely cannot have been these central and long-trumpeted features of the early theory that turned him against it.

About the arguments so far considered Russell is explicit that they are “not decisive” and do not show the early theory to be “logically impossible.” Yet he introduced the discussion by remarking that “the possibility of false judgements *compels us* to adopt” the multiple-relation theory (1910: 118; emphasis added). Only one passage even aspires to the demonstrative status this advertisement leads us to expect. In it Russell argues as follows:

If we hold that

- (i) the judgment that Charles I died on the scaffold has as its objective “the actual event” of Charles I’s death on the scaffold,
then we must in parallel hold that
- (ii) the judgment that Charles I died in his bed has as its objective the actual event of Charles I’s death in his bed.

But there is no such event, “since to say that there ever was such a thing as ‘Charles I’s death in his bed’ is merely another way of saying that Charles I died in his bed.”

So not (ii).

So not (i).

One asks: what have *events* to do with the early theory? We saw that the early theory can admit the notion of a corresponding entity, designated by a “verbal noun,” and existing only when the judgment is true. Those entities might be identified with events. This identification, however, immediately cancels any commitment to (ii); and in any case, those entities had only a peripheral and optional standing in the early theory. Perhaps, then, “events” are the early theory’s *facts*. But this identification undermines the move from (i) to (ii) in a different way. The judgment that Charles I died on the scaffold, according to the early theory, has as its objective a proposition which, because it is true, is describable as the fact that he so died. The parallel between true and false judgments requires that the judgment that Charles I died in his bed also has a proposition as its objective, but it clearly does not require that this proposition be describable as a fact.⁶ Thus it seems that on no interpretation of its terms will the above argument connect with the early theory. Yet it is clearly this argument on which Russell later relies in declaring the early theory to be “impossible” (1912: 193–4).

The reasoning just gone through shows that the early theory’s propositions must be distinguished from what Russell now calls “events.” It reproduces his own earlier reasoning, about complex entities denoted by such phrases as “the execution of Charles I” or “the blackness of the table,” which led him to conclude that “the proposition must be somehow distinguishable from such complexes.” But this conclusion, Russell immediately conceded, makes it “very difficult to see what the proposition is” (1904: 471). The real explanation for Russell’s change of mind must be whatever turned this *difficulty* into an *impossibility*, with the effect that reasoning he had all along endorsed now seemed to yield the conclusion that a proposition must be distinct from anything there actually is—that is, that there can be no such things.

The relevant considerations were more clearly set out in *Theory of Knowledge*, when Russell asked “what kind of entity a false proposition could be” (1913: 110). Recall from our statement of the early theory the proposition that Andrew loves Beatrice. This, we said, has as its constituents the objective entities it concerns, Andrew, Beatrice, and the relation of loving. Its identity depends not only on these constituents, but on how they are configured in it. For instance, it is distinguished from the proposition that Beatrice loves Andrew in that the relation of loving occurs in it as relating Andrew to Beatrice rather than vice versa. But is this not to say that the relation of loving *does* relate Andrew to Beatrice? And is this not to say that Andrews loves Beatrice—that Andrew, loving, and Beatrice are so combined as to constitute the *fact* that Andrew loves Beatrice? In short: try to specify what a false proposition would be, and you will find that what you

⁶ Cartwright (1987: 77–8) makes this same point, against an argument he attributes to Moore.

have specified cannot fail to be true. “It seems,” Russell concluded, “that nothing which is actually composed of these objects is the proposition; and it is not credible that anything further enters into the proposition” (1913: 110). It is, we think, this problem that Russell has in mind in 1910 in writing that his early theory “leaves the difference between truth and falsehood quite inexplicable” (1910: 119). What is “inexplicable” on the early theory is not merely *what* the difference consists in, but *how there can be* any difference.⁷

It is in retrospect rather difficult to see how Russell had held out for so long against this conclusion. Already in the *Principles* he attributed the unity of a proposition to a relation’s occurring in it as “actually relating,” saying that “[the relation of loving] which occurs in the proposition actually relates A and B” (1937 [1903]: section 54). We quoted similar statements above, including the summary that a proposition consists “solely in the fact of relatedness” (1904: 437), which should surely have sounded a warning.⁸ But the case is not really one of a point that Russell had simply failed to see. More plausibly, it is a point that he gradually lost reason to think it *must* be possible to evade. In the passage just referred to, and throughout his exposition of the early theory, Russell struggled to characterize something he thought there must be, a “special and apparently indefinable kind of unity” which is distinctive of propositions and can ground the duality of truth and falsehood. By 1910 he had come to accept that the struggle must end in failure, and that the project was misconceived.

6.2.3 The character of Russell’s new theory

Our contention in the previous section was that the problem of false propositions—the impossibility of specifying what a false proposition could be—was the decisive consideration in motivating Russell’s new theory. In this section we sketch the perspective on the new theory that is encouraged by this contention.

As we saw, Russell’s other complaints against his early theory emphasize its failure to accommodate the intuition that drives a correspondence theory of truth. This is, in essence, that a true judgment is *made true* by the existence of some entity corresponding to it. Or, in something closer to Russell’s formulation, it is that for any true judgment there exists a corresponding entity which would not exist if the judgment were false, and that the distinction between truth and falsity is to be explained by the existence or non-existence of this corresponding entity. If we suppose that the desire to accommodate this intuition was central in motivating Russell’s change of mind, this will encourage us to regard the correspondence theory of truth as an intrinsic part of the view he was led to: we will hold, in effect, that Russell’s new theory of judgment is designed to suit a

⁷ This account of Russell’s rejection of propositions was, we think, first clearly laid out by Richard Cartwright (1987: 82–4).

⁸ For further illustration, consider the joint effect of “a proposition consists of a relation between its terms” (1905: 495) and “objects in relation form a complex object, which may be called a fact” (1906–7: 45).

correspondence theory of truth. A better view, we suggest, is that Russell's correspondence theory of truth is designed to make good a deficiency in his new theory of judgment, one that is owed to the general style of ontological theorizing within which this theory is developed.

The problem of false propositions poses a general challenge to Russell's early ontology. Responding to it would call for thoroughgoing revision to all of his analyses or constructions that exploit the general category of propositions, conceived as complex entities that might be true (and, if so, facts) or false (and, if so, fictions). For instance, when he wrote his inconclusive 1906–7 paper, Russell had recently developed a type-theoretic solution to the paradoxes in the foundations of mathematics through his “substitution theory” (Russell 1973 [1906]). The basic idea of this theory is that the only genuine, ontologically-committing quantifications are over propositions and individuals. Without propositions a quite different foundation for Russell's type-theoretic logicism would be needed. Clearly, then, the problem was anything but local. In his essays on judgment and truth Russell tended not to emphasize the more “technical” issues in which commitment to propositions is at stake. Even so, in “On the Nature of Truth” his first argument against the proto-multiple-relation theory and in defense of propositions is set on general logical ground, holding that false propositions are needed for the analysis of truth-functionally complex statements (1906–7: 48). And it is interesting to note that when, in *Theory of Knowledge*, Russell reviews the debate from the perspective of his later views, he declares that this general logical argument provided the *only* ground for his early commitment to propositions (1913: 153).

A similarly general strategy⁹ for responding to this argument is, though, already outlined in the 1906–7 paper: “a valid analysis, we shall have to contend, must break up any apparent subordinate complexes into their constituents” (1906–7: 48). That is to say—taking $\langle aRb \rangle$ as our representative “subordinate complex” or proposition—that anything of the apparent form $\varphi(\langle aRb \rangle)$ must be recast as being of the form $\Psi(a, R, b)$. This strategy provides the analytical substance of the doctrine that propositions are “incomplete symbols”—that analytical reconstrual will reveal that phrases taken to express propositions have no unitary meaning. Russell's new multiple-relation theory conforms to this general strategy. It says that the fact of someone's making a judgment consists in a relation the subject bears only to the several constituents of the erstwhile proposition— $\Psi(a, R, b)$ —and not to any complex entity forged out of those constituents— $\varphi(\langle aRb \rangle)$. But so far as the 1910 statement of the theory takes us, and insofar as it merely instances the general strategy, this is *all* that it says. Hence Russell has to add to this rump of a

⁹ Only “similarly” general, since this strategy would not be invoked in *every* issue involving propositions. For instance, the substitution theory's quantification over propositions is replaced in *Principia* by the acceptance unreduced of quantification over propositional functions; see Bostock (2012: 205–6). And the suggestion that truth-functionally complex propositions such as “ p or q ” should be structurally recast in this fashion is plainly crazy—though, amusingly enough, there is evidence that Russell set Wittgenstein to work on the suggestion in 1912 and that Wittgenstein pursued it enthusiastically; see Wittgenstein (2008: 32); Potter 2009: 51–2).

theory, to transform it into something recognizably a theory of *judgment*, an account of what it is for a judgment so formed to be true: this will be so just in case the various objects to which the subject is related in making the judgment are themselves so related as to constitute a second complex entity, a corresponding fact.

Viewed from the perspective we have reached, then, neither the problem Russell faced nor the immediate resolution he proposed for it has specifically to do with judgment or “the analysis of cognition” (Ramsey 1927: 142). This perspective allows us to recognize Russell’s new account of judgment and its associated theory of truth as separable components in a two-step approach. It also highlights a real oddity in the division of labor between these two steps. One minimally expects of a theory of judgment that it will seek to explain judgment’s centrally defining feature: that a judgment is to the effect that something is the case, or that it has content. Russell’s first step does not do this. His second step is needed because the first achieves so little.

This was Ramsey’s diagnosis. He began “Facts and Propositions” by first outlining and then importantly supplementing Russell’s 1910 arguments for the conclusion that “a judgement has not one object but many, to which the mental factor is multiply related” (1927: 142). “But,” Ramsey continued,

to leave it at that, as he [Russell] did, cannot be regarded as satisfactory. There is no reason to suppose the multiple relation simple . . . and it is desirable that we should try to find out more about it . . . Similarly, a theory of descriptions that contented itself with observing that “The King of France is wise” could be regarded as asserting a possibly complex multiple relation between kingship, France, and wisdom, would be miserably inferior to Mr Russell’s theory, which explains exactly what relation it is. (ibid.)

The comparison with the theory of descriptions, the archetype of Russell’s theories of “incomplete symbols,” shows that Ramsey’s complaint is directed toward what we have called Russell’s first step: it is the account of judgment itself that is “miserably inferior” to what we should expect. The continuation makes plain Ramsey’s view that, if this deficiency were rectified, Russell’s second step would no longer be needed.

Our perspective also makes clear that the form taken by Russell’s second step, his correspondence theory of truth, is determined less by the specifics of the first step than by the general style of ontological theorizing that generates both the initial problem and the schema for resolving it. Speaking very broadly, Russell’s analytical approach asks, of any given phenomenon, what kind of configuration, of what kinds of objects, might constitute it. His early commitment to propositions sought to reserve within this approach a special and central domain that would be governed by the demands of an analysis of cognition—a domain manifesting a “special” kind of complexity, not needed in the analysis of any merely external phenomenon, that could internally ground the duality of truth and falsehood. But the tendency of the general approach was always against this, and in succumbing to the problem of false propositions Russell surrendered this special domain. Thus the task of his new theory of judgment becomes that of subsuming

representation within a general ontological framework whose basic categories and contours are not dictated by its requirements. Had it been successful, its achievement would have been, in one sense of a much-used phrase, to “explain thought from outside,” as being merely one amongst various configurations of what there is.

The repudiation of this ambition is the basic ground of Wittgenstein’s criticism of Russell’s theory, presented in conversation and in letters in the late spring and summer of 1913. The exact analysis of Wittgenstein’s criticism has been endlessly discussed in recent years, and we will not here add to that discussion. For our purposes, the relevant point is only that the central premise of Wittgenstein’s complaint, like Ramsey’s, targets the inadequacy of Russell’s first step. It must, he insists, be made plain in the account of judgment itself, and cannot be left to any supplementary theory to explain, that judgment is to the effect that something is the case, that it has content, and is true or false (Wittgenstein 2008: 40).

Wittgenstein’s final comment on the matter to Russell that summer was that his central objection “can only be removed by a correct theory of propositions” (2008: 42). This theory is our topic in the following section. We will show, as Ramsey did, that it displaces Russell’s second step too.

6.3 WITTGENSTEIN

We have outlined the route by which Russell arrived at his 1910 view that, “when we judge truly some entity ‘corresponding’ in some way to our judgment is to be found outside our judgment, while when we judge falsely there is no such ‘corresponding’ entity” (1910: 119). We move on now to consider what Wittgenstein made of this idea.

We should enter a terminological note at the outset. “Proposition” is the accepted translation of Wittgenstein’s “*Satz*,” a meaningful sentence. In this section we follow this usage, since for the most part the objective, abstract entities that Russell’s early theory called “propositions” are no longer in question. Henceforth, then, a “proposition” will generally be the expression of a judgment, not its supposed “objective,” and any exceptions to this will be explicitly flagged.

6.3.1 The supposed role of a correspondence theory in the *Tractatus*

Russell offers the following example of his correspondence theory:

If A loves B, there is such a complex object as “A’s love for B”, and vice versa; thus the existence of this complex object [which Russell identifies as a fact] gives the condition for the truth of the judgment “A loves B”. (1910: 123)

Numerous formulations in the *Tractatus* suggest that Wittgenstein shared the same basic conception. One central instance is the following:

If the elementary proposition is true, the atomic fact exists: if the elementary proposition is false, the atomic fact does not exist. (TLP 4.25)

It was for many years an orthodoxy amongst commentators on the *Tractatus*, and it seems still to be the dominant view amongst less specialist readers, that these formulations are to be taken at face value: Wittgenstein, like Russell, offers a correspondence theory of truth; that is, like Russell, he holds that the obtaining of a truth condition consists in the existence of a corresponding fact. We will argue that this widespread view is mistaken. But it is certainly not groundless. After all, Wittgenstein wrote to Russell in 1919, in explanation of his terminology, that “Sachverhalt [atomic fact] is, what corresponds to an Elementarsatz [elementary proposition] if it is true” (2008: 98). And the exposition Russell provides in his “Introduction” to the *Tractatus*, an account informed by a week’s intense discussion with Wittgenstein at the end of that year, signals no major difference between them on this point.¹⁰ However, a more substantive reason for attributing a correspondence theory to the *Tractatus* is that it appears to explain distinctive and central features of Wittgenstein’s account, which are naturally viewed as responses to problems faced by a correspondence theory.

The first of these is what we will call “the problem of logical complexity.” The passages just cited from Russell and Wittgenstein both have to do directly with *atomic* propositions (or judgments). How should we understand the account as extending to explain the truth of logically complex propositions? A straightforward implementation of the correspondence idea would have it that the proposition “ p or q ” is true if and only if the “disjunctive fact,” p -or- q , exists. But how should this fact be conceived of? It cannot consist of the facts p and q related by a “disjunction relation,” for then the existence of the fact p -or- q would entail the existence of the fact p , and so the truth of the proposition “ p or q ” would entail the truth of the proposition “ p .” And what about “negative facts”? How should we conceive the fact *not*- p , such that it will exist just in case the fact p does not? There are in fact two problems here. The first runs parallel to the problem over disjunctive facts: if the fact *not*- p were some sort of compound entity, with the fact p as a part, then the truth of “not p ” would entail the *truth* of “ p ” rather than, as we should expect, its falsity. But suppose this first problem can be overcome, so that the facts *not*- p and p are understood as distinct facts, neither including the other; we will still lack any explanation for why one of these should exist precisely when the other does not.

¹⁰ Russell *does* present Wittgenstein as holding that the statement that a certain complex exists “reduces to” the statement that its constituents are related in a certain way (Russell 1922: 12). But the intended scope of this remark is not altogether clear—does it apply only to Wittgenstein’s concerns at TLP 3.24?—and it does not dissuade Russell from the summary that “facts are what make propositions true” (1922: 11).

At TLP 4.0132 Wittgenstein wrote, “My fundamental thought is that the ‘logical constants’ do not represent.” One common understanding of this thought connects it with his early insistence that “whatever corresponds in reality to compound propositions must not be more than what corresponds to their several atomic propositions” (Wittgenstein 1913: 98). For instance, the disjunctive proposition “ p or q ” will have no worldly correlate beyond those of the atomic propositions “ p ” and “ q ,” hence no worldly correlate of which the meaning of the connective “or” might be a constituent. Thus the intention of the fundamental thought, on this understanding of it, is to deny that a correspondence theory is to be extended in the simple way we imagined to non-atomic propositions. There is no disjunctive fact, p -or- q , whose existence constitutes its being the case that p or q . Rather, the only facts envisaged in Wittgenstein’s account are those corresponding to the atomic propositions “ p ” and “ q ,” and its being the case that p or q may be constituted either by existence of the fact p , or by the existence of the fact q , or by the existence of both. Similarly, Wittgenstein countenances no such entity as the negative fact *not- p* , guaranteed somehow to exist whenever the fact p does not. Rather, its being the case that not p —that truth condition obtaining—will consist in the *non*-existence of the fact p .

The second problem for the correspondence idea, which we will call “the problem of falsity,” appears in Wittgenstein’s *Blue Book*:

How can one think what is not the case? If I think that King’s College is on fire when it is not on fire, the fact of its being on fire does not exist. Then how can I think it? (Wittgenstein 1958: 31)

For a thought to have as the condition of its truth the existence of the fact that King’s College is on fire is for it to be connected with that particular fact. (A thinker of such a thought might be said to think *just that fact*.) But how then is falsity possible? How can a thought be connected with a particular fact if no such fact exists to be connected with?¹¹

In section 6.2 we explained how Russell’s new account of the constitution of a judgment responds to this problem. A fact, Russell maintains, is a complex; and a thought is a relation between the thinker and the several, separate components of the complex whose existence would constitute the obtaining of its truth condition. Thus a thought may connect in the relevant way with a particular, non-existent fact by virtue of connecting to that fact’s existing components. In outline, then, Russell’s answer to the problem of falsity is the one envisaged by Wittgenstein in the continuation of the passage from the *Blue Book*:

“How can we imagine what does not exist?” The answer seems to be: “If we do, we imagine non-existing combinations of existing elements”. (1958: 31)

¹¹ Obviously this is connected with what section 6.2.2 called “the problem of false propositions.” They have in common the reasoning that falsity is impossible if (a) the content of a judgment is given by its association with a unitary, complex entity, and (b) the only such entities are facts. But we have introduced a different label here to indicate that the problems arise at different points in the dialectic. The argument of section 6.2.2 turns (b) against a position, Russell’s early theory, that is clearly committed to (a). The problem of falsity, as intuitively sketched here, arises in advance of any such clear theoretical commitment.

This same answer has been attributed to the *Tractatus*. A Tractarian atomic proposition pictures, or presents, a state of affairs: the sense of the proposition *is* the state of affairs it presents (TLP 2.221, 4.031). Hence to understand the proposition, or to have the thought expressed by it, is to “know [*kennen*] the state of affairs presented by it” (4.021). Here one might sense the problem of falsity looming. But the proposition is *articulate* (3.141), consisting of names in combination (4.22); and the connection with reality in virtue of which the proposition presents the state of affairs it does—what Wittgenstein calls the “representing relation”—consists only in coordinations between these names and objects, constituents of the presented state of affairs (2.1513–14). Hence one understands the proposition, and knows the state of affairs it presents, provided only that “one understands its constituent parts” (4.024; cf. 3.4); and so one can understand the proposition without knowing whether it is true (4.024). These claims, one might suggest, constitute Wittgenstein’s response to the problem of falsity. For Wittgenstein, as for Russell, a thought or judgment may have a truth condition whose obtaining consists in the existence of a particular, non-existent fact by virtue of relations only to the several existing entities which, suitably configured, would constitute that fact.

A correspondence theory, then, has a central place in familiar interpretations of the *Tractatus*, and on those interpretations it has a centrally motivating place in Wittgenstein’s theorizing. We want now to outline an alternative.

6.3.2 An alternative

The correspondence theorist conceives of facts as things that exist or fail to exist in the world, as things that are present in or absent from the world. For *a* to love *b* is for there to exist in the world such a fact as “*a* loving *b*.” To conceive of facts in this way perhaps need not be to think of them as logically of a piece with the ordinary objects they concern, in all the ways Russell suggests when he talks of facts as “go[ing] about the real world,” or as “exist[ing] in the actual world of nature” (1918: 196–7), or again, when he insists that “just as much as particular tables and chairs, [they] are part of the real world” (1918: 164). But it *is* to think of them as in some way “portions” or “chunks of reality” (cf. Strawson 1950: 211), and a “chunk” conception of facts is far from obligatory.

Wittgenstein points toward an alternative in another later passage that raises the intuitive problem of falsity:

It makes no sense to say “I am killing something that does not exist”. I can imagine a stag that is not there, in this meadow, but not kill one that is not there. And “to imagine a stag in this meadow” means to imagine *that* a stag is there. But to kill a stag does not mean to kill *that* . . . (Wittgenstein 1974: 137; emphasis in original)

A fact, Wittgenstein here suggests, is nothing like a stag—it is no kind of chunk of reality—for then it would be as impossible “to think what is not” as it is to kill something there is not. Rather, a fact is always *that p*: a fact is a way things are, and “to think what is not” is to think *that* things are a certain way when in fact they are not that way.

This position bears some spelling out. On the correspondence idea, a fact was something in terms of which the obtaining of a truth condition was *explained*: the obtaining of a truth condition, in the basic case at least, consisted in the existence of a fact. On the present suggestion, however, a fact simply is a truth condition that obtains. A fact is not a chunk whose existence constitutes the obtaining of one truth condition and whose non-existence constitutes, perhaps, the obtaining of another. Rather, a fact is itself the satisfaction of a *single* truth condition: the proposition “*p*” has as its truth condition *that p*.

How should we understand this identification of fact and truth condition? Well, it is not merely terminological. The claim is not that Russell (say) should use the word “fact” not for his complexes but rather for truth conditions. Nor, though, do we have here a substantial positive thesis. What we have, rather, is the substantial negative thesis that the obtaining of a truth condition is not to be understood in terms of a distinct theoretical item worthy of the name “fact.” In particular, the obtaining of a truth condition is not in general to be understood in terms of the *existence* of an item of a distinctive theoretical kind. Rather, the word “fact” has nothing to apply to other than, simply, truth conditions—and of course we may add, since the grammar of the English “fact” is such that to talk of the fact that *p* is, in most contexts, to commit oneself to its being the case that *p*, that the word “fact” will typically apply simply to those truth conditions that obtain.

We can consolidate our understanding of this thought in a couple of straightforward ways. First, we can note that the present proposal will respond to the correspondence theorist’s emphasized talk of fact *existence* by—as it were—identifying a fact’s existence with the fact itself. To recognize, affirm, deny, or hypothesize the existence of a certain fact will simply be to recognize, affirm, deny, or hypothesize that fact. Where *A* is any kind of chunk of reality, to identify the existence of *A* with *A* itself would be an incongruous grammatical blunder. As Ramsey remarked, no one will confuse the existence of a King of Italy, the fact that Italy has a King, with the King himself (1927: 141). Similarly, if an event, such as a football match, is a portion of reality, then no one will confuse the event itself with the existence of this event or the fact that it took place. By contrast, to distinguish the existence of a fact from the fact itself is mere long-windedness. To say that a certain fact exists, or obtains, is just to say of a certain way things may be that it exists, or obtains, and this will mean nothing other than that things are indeed that way. As Ramsey wrote, “‘The fact that *a* has *R* to *b* exists’ is no different from ‘*a* has *R* to *b*’” (1927: 143).

Secondly, we can observe that, from the current perspective, the problems of falsity and logical complexity simply disappear. Or rather, they don’t so much as appear. The problem of falsity—the would-be worry of how I can think the fact *p* when that fact does not exist—does not get beyond the question: how can I think that *p* when it is not the case that *p*? That is to say, it does not get beyond the question of what it is for a thought to be a thought *that p*, i.e. of what it is for a thought to have a particular truth condition. This is of course an important question, but it is not one concerned in any special way with *falsity*. More, identifying facts with truth conditions that obtain removes both the need and the platform for worrying about negative and disjunctive facts. If things may

be such that q , and may be such that r , then *of course* a further way that things may be, distinct from either of those, is such that q or r . And trivially, things will indeed be this way, and this distinct fact will exist, whenever the fact that q exists or the fact that r exists. For a fact that p to exist just is for it to be the case that p , and if it is the case either that q or that r , then it is the case that q or r .¹² As for negative facts, well, if things may be such that q , then *naturally* things may also be such that not q , and *naturally* the latter fact will exist just in case the former does not. (Every truth condition has an opposite: for every way things may be there is an opposite way things may be such that for things to be the one way is for them not to be the opposite way.)

With the alternative so far sketched we can raise the question, what is the Tractarian position on these matters? We acknowledged that Wittgenstein's talk of fact existence, scattered liberally through the *Tractatus*, can suggest that he has a chunk conception of facts, and more particularly that he exploits this conception in offering a correspondence theory of truth. And we saw that this understanding is further supported by the ready interpretation it offers of Wittgenstein's "fundamental thought" that the logical constants do not represent: in the assumed context of a correspondence theory, this thought will be understood as a rejection of logically complex fact-chunks involving such things as a relation of disjunction or a property of negation. Counting against this, however, and in favor of attributing to the *Tractatus* the alternative just sketched, is that whenever Wittgenstein wants to emphasize that something is a fact he invariably lays stress on the idiom "*that p*." So for instance we find:

A propositional sign is a fact . . .

Not: "The complex sign ' aRb ' says ' a stands in relation R to b '"; but rather: "*That* ' a ' stands in a certain relation to ' b ' says *that* aRb ". (TLP 3.14, 3.1432; Wittgenstein's emphases)

The contrast emphasized here makes plain that, when Wittgenstein insists that "only facts can express a sense" (3.142), his primary intention is to deny that a *complex* may do so. This denial would hardly be intelligible if Wittgenstein himself conceived of a fact as any kind of complex.

If such reasoning persuades us that the alternative view is indeed Wittgenstein's, then we will of course have to account for the contrary evidence. But at least the first step in this task is easy. It is a *part* of the alternative view that talk of fact existence is, in one sense, cheap: it carries no theoretical weight. And, given the inveterate nominalizing tendency of ordinary grammar (cf. Strawson 1950: 197–8), avoiding such talk is, in

¹² We should perhaps note, to prevent a misunderstanding, that to be untroubled by disjunctive facts is not to countenance *irreducibly* disjunctive facts. By reasoning exactly like that in the main text, we can say that, if it is the case that p or q (i.e. if the fact that p or q exists), then it is the case either that p or that q (i.e. either the fact that p exists, or else the fact that q exists). Nothing in the alternative view we are recommending contradicts Wittgenstein's theory that the truth or falsity of *every* proposition is determined by the truth or falsity of atomic propositions: if it did, the view would be hopeless.

another sense, costly: our exposition of the alternative view provides, we fear, enough illustration of the prolixity needed to counter this tendency. (If not, then private experiments in recasting some of Wittgenstein's most prominent examples of "fact-talk," in the 4.3s and 4.4s, will prove the point.) Wittgenstein's taste for succinctness is thus a more than adequate explanation for his allowing formulations suggestive of a chunk conception of facts to stand.

But then how is the issue to be decided? We have taken Russell's adherence to a chunk conception as clear, and have treated such passages as the following as clear examples of his invoking this conception in the service of a correspondence theory of truth:

The judgment that two terms have a certain relation *R* is a relation of the mind to the two terms and the relation *R* with the appropriate sense: the "corresponding" complex consists of the two terms related by the relation *R* with the same sense. The judgment is true when there is such a complex, and false when there is not. (1910: 124)

Yet it seems that Russell's immediate purpose would be just as well served by a definition that ran instead:

The judgment is a relation of the mind to the two terms and the relation *R* with the appropriate sense; it is true when those two terms are related by the relation *R* with the same sense, and false when they are not.

This alternative definition inflates into Russell's just when we add the claim that for *a* to stand in relation *R* to *b* is for the complex "*a* in relation *R* to *b*" to exist. But why should we regard *this* claim as a substantive thesis, rather than, as the alternative view outlined in this section would construe it, a mere tautology? We suggest that this question offers the clearest route to resolving the broader issue of the position of the *Tractatus*. In the following section we will consider two very different answers to it—two different ways in which talk of fact existence might be made substantive—one grounded in Russell's views, and one that has been supposed to represent Wittgenstein's; and we will show that neither can be attributed to the *Tractatus*.

6.3.3 The substance of talk about facts

Russell specifies as the condition for the truth of an atomic judgment the existence of a corresponding complex or fact. But what gives substance to this specification, and to the notion of a complex that he invokes in it, cannot be seen in this application of it alone. Instead, we must attend to wider theoretical context on which the specification draws. For Russell, complexes are not *merely* the realizations of truth conditions. He conceives of them as entering into other kinds of relations, and thus as serving other theoretical purposes. Most importantly, he holds that complexes figure *as objects* as constituents of further complexes, a structuring that will yield substantive explanations of certain

necessary relations amongst truth conditions. It is such further applications of Russell's general theory of complexes, and not in itself his theory of judgment—not in itself, i.e. his theory of a judgment's possession of a truth condition—that provide substance to his correspondence idea.

The example given prominence in "On the Nature of Truth and Falsehood" concerns perception. Russell there holds that the complex "knife-to-the-left-of-book"—the complex whose existence constitutes the knife's being to the left of the book—is a constituent, along with Jack and the relation of perceiving, of the further complex "Jack-perceiving-<knife-to-the-left-of-book>," this latter complex again existing just in case Jack enjoys the relevant perception. Since the existence of a whole requires the existence of its parts, Russell has here an explanation of why Jack's perception is possible only if the knife is indeed to the left of the book, and more generally, of why perception is "infallible" (1910: 122–3).

Turning with this to Wittgenstein, we can straightaway note that he emphatically rules out any such context for thinking of facts as chunks. He both vigorously disassociates facts from objects and insists that objects are simple. A fact is not a possible constituent, as an object, of further facts. As for perception, Wittgenstein writes,

To perceive a complex means to perceive that its constituents are related to one another in such and such a way. (TLP 5.5423)

Perceiving "*a* in relation *R* to *b*" is to be understood as perceiving *that a* is in relation *R* to *b*. And perceiving that *aRb*, Wittgenstein further implies, is a matter of having a true perceptual representation that *aRb*.

We noted Russell's explanation of the "infallibility" of perception: as a relation between a mind and a complex, a perception is possible only if the complex exists. Russell is, however, quick to point out that this does not confer infallibility on any perceptual *judgment*: while in perception the mind is acquainted with the complex as a single object, the judging mind has before it only the several constituents of that object. The transition between the two—e.g. from the awareness of "this-before-that" to the judgment "this is before that"—Russell describes as "an analysis," and "we should not understand the analysis if we were not acquainted with the meanings of the terms employed" (1911: 150). This understanding, Russell however holds, must rest on a prior such transition, whereby the meanings of terms are "abstracted" from the wholes in which they are first presented:

When we see a white patch, we are acquainted, in the first instance, with the particular patch; but by seeing many white patches, we easily learn to abstract the whiteness which they all have in common, and in learning to do this we are learning to be acquainted with whiteness. (1912: 158–9; cf. 1911: 150)

Thus the representation of a complex *as a complex* rests on acquaintance with its constituents, which in turn rests on awareness of such complexes *as objects*. In this way

the multiplicity of relations into which complexes may enter plays an essential role in Russell's account of thought.

Turning to the *Tractatus* one again sees that any such dual role for facts is firmly rejected:

The meanings of primitive signs can be explained by elucidations. Elucidations are propositions which contain the primitive signs. They can, therefore, only be understood when the meanings of these signs are already known. (TLP 3.263)

Wittgenstein here adds no commentary even to soften the circularity that Russell aimed to avoid. Instead, it is left starkly signaling the commitment that entails its inevitability: there is, Wittgenstein clearly implies, no “pre-propositional” presentation of facts; rather, a fact is simply what a proposition presents.

Interestingly, Wittgenstein at the same time signals a distance from his former self. The *Notes on Logic* include a curious—and in the end, we think, an incoherent—passage explaining how meaning is conferred on a simple predicate (Wittgenstein 1913: 104). The sound thought that Wittgenstein will retain from this account is that to lay down the meaning of a predicate is to establish a rule determining the truth conditions of basic sentences containing it. But he imagines that this will be done by (somehow) “dividing . . . the facts” into those “of like sense” and those “of opposite sense” with the intended meaning. However exactly this is supposed to work, it seems to presuppose that the facts are available to and arrayed in front of a subject *in advance* of his understanding propositions expressing them—just as they are on Russell's abstractionist account. Now it is clear that in 1912 Wittgenstein was working entirely within Russell's theory of complexes.¹³ What we called the sound point in this passage from the *Notes* was already a departure from Russell.¹⁴ But the confused ideas surrounding it—not to mention Wittgenstein's tortuously obscure formulation of those ideas—indicate that this was only a first step, and that he had not fully broken free of Russell's framework. What matters most to our case, though, is that there is no trace whatever of those ideas when the same issue is reached in the *Tractatus*.

It seems, then, that we will need to look elsewhere if we want to think of Tractarian facts other than as truth conditions that obtain. But where? Well, the obvious place is where, for Russell, we didn't find the relevant provision: namely, in the theory of judgment itself. Thus the thought we explore in the second part of this section is this: perhaps Wittgenstein's theory of the *possession* of a truth-condition will carry substantial implications for what it is for a truth condition to obtain.

The broad structure of Russell's new theory is that a judgment connects to a fact *in virtue of* having a truth condition. Although section 6.2.3 criticized his implementation of

¹³ Cf. fn. 9.

¹⁴ Russell was persuaded of the point by the *Notes*, and gave it prominence in his “Logical Atomism” lectures (1918: 182).

this thought, it is clearly the intention of Russell's theory that a judgment's possession of a truth condition should be explained by its connections only to a multiplicity of objects. Unitary, complex entities, facts, are to enter the account only secondarily, in explanation of what it is for a judgment's truth condition to obtain; and we have argued that, while this second step does involve a chunk conception of facts, it does so for reasons external to the theory of judgment. The broad structure of a view opposite to Russell's would suggest instead that a judgment or proposition has a truth condition *by virtue of* connecting to a fact—or perhaps to some unitary, fact-like entity that we might more neutrally call a “state of affairs.” According to this anti-Russellian suggestion a proposition's first achievement is to “pick out” a state of affairs; subsequently, it does something like “asserting” this state of affairs; and it is only through this second step of “asserting” that the proposition comes to have a “sense” or truth condition. If a theory of this general shape of the functioning of an atomic proposition could be ascribed to Wittgenstein, then all would be set for thinking of Tractarian facts other than as ways things are. An “atomic fact” or state of affairs (*Sachverhalt*) would be the object of a “picking out” relation, and—depending on how the assertion part of the theory is understood—the door would be open for thinking of the obtaining of an atomic truth condition in terms of the *existence* (or something similar) of such an object.

Many commentators have understood the *Tractatus* as proposing just such a two-step theory of atomic propositions.¹⁵ First, the proposition picks out a state of affairs—it *depicts* or *images* a state of affairs in a way which does not presume the existence of that state of affairs. Subsequently, the proposition asserts the picked out state of affairs, coming thereby to have a truth condition. There is, however, ample reason to reject any such understanding of Wittgenstein's picture theory. To begin with an indirect reason, the attribution is seriously undermined by criticisms Wittgenstein makes of Frege:

The verb of a proposition is not “is true” or “is false”, as Frege thought: rather, that which “is true” must already contain the verb [i.e. it must already have a truth condition].

Every proposition must *already* have a sense [again, a truth condition]: it cannot be given a sense by affirmation. (TLP 4.063–4.064; emphasis in original)

Rightly or wrongly (wrongly, we think), Wittgenstein here *accuses* Frege of explaining the possession of a truth condition by reference to an affirmation, or asserting as true, of some prior item. Moreover, his criticism does not target—it hardly even mentions—the detailed features of Frege's theory that (allegedly) commit him to this account: it is the two-step structure *itself* that Wittgenstein objects to. This makes it at least unlikely that Wittgenstein himself adhered to a theory with the same structure.

However, we need not rely on such indirect considerations. What most severely undermines the attribution to Wittgenstein of a two-step picture of atomic judgment

¹⁵ The most recent example we know of is Hanks (2012).

is the complete and conspicuous absence of any such idea from Wittgenstein's central statement of his theory:

In the picture the elements of the picture are the representatives of objects.

A picture is a fact.

That the elements of the picture are combined with one another in a definite way, represents that the things are so combined with one another. (TLP 2.131, 2.141–2.15)

A proposition's having a certain truth condition—its representing *that something is the case*—is here explained *directly* in terms of its elements referring to objects; it is not explained via the proposition as a whole imaging some non-truth-condition fact-chunk.

In this section we have considered two routes by which talk of the existence of facts might be made substantial, or inflated to carry theoretical weight. The first, Russellian route locates the substance of this talk in the account that is offered of *the obtaining* of a truth condition. That a judgment *has* a truth condition is explained by reference to connections made only to a multiplicity of objects: the judgment does not possess a truth condition by virtue of relating, directly or indirectly, to a single, unitary entity. Such entities are invoked only in explaining what it is for a judgment's truth condition to obtain. But what gives substance to the conception then invoked, we argued, are the further theoretical applications Russell makes of his general theory of complexes, applications that Wittgenstein clearly rejects. The second, anti-Russellian route is represented by the "two-step Wittgenstein" just considered, and locates the substance of fact talk in the account of a judgment's *possession* of a truth condition. This "two-step Wittgenstein"—or at least, the most exegetically plausible version of this character—holds with Russell to the correspondence theory of truth. Against Russell, however, he does not offer a multiple-relation theory of judgment: he denies the core thought of a multiple-relation theory, that a judgment's connections to several objects are sufficient for its possessing a truth condition, holding instead that these connections provide only for it neutrally to image a single, possibly non-existent fact, which then in turn provides, through the second step of asserting the imaged fact, for the judgment's possession of a truth condition. But this "two-step Wittgenstein" is not the real Wittgenstein. The real Wittgenstein does, like Russell, offer a multiple-relation theory of judgment. Interestingly, though, whilst Russell's multiple-relation theory of judgment is a response in part to his need to address the problem of falsity arising from his commitment to a correspondence theory of truth, Wittgenstein's offering a multiple-relation theory of judgment is an aspect of his *not* offering a correspondence theory of truth, and so *not* being confronted by the problem of falsity.

Nothing in Wittgenstein's theorizing, we conclude, provides for interpreting those points at which he talks of fact existence as the expression of a correspondence theory, or indeed of any theory in which facts are distinguished from truth conditions that obtain. Nothing in the *Tractatus* provides for thinking of the sentence "the fact '*aRb*' exists" as anything other than a periphrasis of the sentence "*aRb*."

6.3.4 Facts as incomplete symbols

What we have argued at length in this section, Ramsey appears to have seen immediately and without effort. In his “Critical Notice,” and in the course of a more substantive discussion to which this observation is almost an aside, Ramsey reports as something Wittgenstein “clearly believes” that

we cannot talk about what makes a fact a fact, nor ultimately *about* facts at all, because every statement apparently about facts is really about their constituents. (1923: 273)

In other words, what Russell had come to believe about his early theory’s abstract “propositions,” in declaring them to be “incomplete symbols,” holds equally, in Wittgenstein’s view, of facts. Neither a “proposition,” conceived as embodying the content of a judgment, nor a “fact,” conceived as the condition whose obtaining constitutes the truth of the judgment, is an entity of any kind, a chunk of reality to which reference might be made. In *both* cases, any apparent reference to such an entity will be exposed by analysis as no more than a misleading form of expression. And indeed, the position Ramsey here ascribes to Wittgenstein is exactly the position Wittgenstein had reached in 1913.

The initially unpromising context for Wittgenstein’s statement of this view is the distinction he drew, in the *Notes on Logic*, between the “sense” and the “meaning” of a proposition. The *sense* of a proposition, as one expects, is its truth condition: “what we understand is the *sense* of the proposition” (1913: 103; emphasis in original), and “to understand a proposition is to know what is the case if it is true” (104). The *meaning* of a proposition, however, is “the fact which actually corresponds to it” (94), and Wittgenstein describes it as “the chief characteristic” of his theory that “*p* has the same *meaning* as not-*p*” (103; emphasis in original). At least as restricted to atomic propositions and their negations, the view seems relatively clear.¹⁶ If an atomic proposition “*p*” is true, then it is the single “positive fact” that *p* which at once grounds the truth of “*p*” and the falsehood of “not *p*”; if, on the other hand, “*p*” is false, then the single “negative fact” that not *p* will ground both the truth of “not *p*” and the falsehood of “*p*.” The view is readily enough understood as a minor modification of Russell’s correspondence theory, one that grants every proposition a worldly correspondent, but without indulging in “fictions” or “objective non-facts.” Thus Wittgenstein’s summary: “*positive* and *negative* facts there are, but not *true* and *false* facts” (97). At any rate, it is clear that Russell himself understood the view in this way (cf. Russell 1918: 185).

Small surprise, then, that Russell was disconcerted to find Wittgenstein also asserting in the *Notes*,

Neither the sense nor the meaning of a proposition is a thing. These words are incomplete symbols. (Wittgenstein 1913: 102)

¹⁶ For an explanation of how the view becomes much less clear when this restriction is lifted, see Potter (2009: ch. 14).

Russell wrote to Wittgenstein, quoting this remark (in its original German, with “*Bedeutung*” for “meaning”) and asking for an explanation:

I understand neither being a *thing*, but I thought the *Bedeutung* was the *fact*, which is surely not indicated by an incomplete symbol? (Wittgenstein 2008: 52)

Wittgenstein first replied,

You say, you thought that *Bedeutung* was the “fact”; this is quite true, but remember that there are no such Things as facts, and that therefore this prop[osition] itself wants analysing! If we speak of “*die Bedeutung*” we seem to be speaking of a Thing with a proper name. Of course the symbol for “a fact” is a prop[osition] and this is *no* incomplete symbol. (2008: 50)

And then, finding this perhaps not entirely clear, he added in a second letter,

The answer is *of course* this: The *Bedeutung* of a prop[osition] is symbolized by the proposition—which is *of course* not an incomplete symbol, *but the word “Bedeutung”* is an incomplete symbol. (2008: 52)

The exchange is instructive, first, in displaying the very different understanding Wittgenstein had reached of what was originally a Russellian thesis; and secondly, in illustrating their contrasting conceptions of analysis. In Russell’s hands the thesis that propositions are incomplete symbols is primarily a reductive ontological claim. The problem of false propositions had persuaded him that there can be no such entities as his early theory had countenanced, and thus that any truths regarding those supposed entities that are to be salvaged must be construed as having to do instead with entities of a different kind. The theory that provides for this reduction is his general theory of complexes. In a word, then, “propositions” are to be reduced to “complexes”; or, translated into Wittgenstein’s words, the “senses of propositions” are to be reduced to the “meanings of propositions.” Now Russell was of course very familiar with the notion that such a reduction need not be ultimate: his own logicist constructions first reduce numbers to classes, then reduce classes in turn to propositional functions, and then those—at least in some versions—to yet more basic entities. So, on Russell’s understanding, it would be open to Wittgenstein to propose that the “facts” (the “meanings of propositions”), to which “propositions” (the “senses of propositions”) are reduced, should themselves be subject to a further reduction. But a move of that kind would of course require the specification of a further reductive class, comprising those entities to which facts are in turn to be reduced; and this is something Wittgenstein signally fails to provide.

The reason is clear from Wittgenstein’s reply. His analytical proposal is targeted, not on the facts that form the meanings of propositions, but on “*the word ‘meaning’*.” That is to say, what Wittgenstein aims to eliminate is not a range of entities, for which some substitute entities would then have to be found, but merely a misleading form of verbal expression responsible for the suggestion that there is any such range of entities to be

reduced. That propositions have sense and meaning is of course not to be denied. What is to be denied is that their having *either* sense *or* meaning is to be understood as their being associated with a kind of “thing,” a chunk of reality to which reference might be made by a “proper name.” Wittgenstein does not, in these short responses to Russell, spell out the analyses by which the conception of the sense or meaning of a proposition as a “thing” is to be avoided. He does, though, clearly indicate their general form. Any expression apparently referring to a proposition’s sense or meaning is to be treated as an incomplete symbol, a form of words to be contextually analyzed by the use of an expression by which that sense or meaning is properly symbolized, that is by the *use* of a proposition. Thus, as a first step, we might have, instead of “‘*p*’ has as its sense the (abstract) proposition *aRb*,” “‘*p*’ says that *aRb*”; and, applying the same model, instead of “‘*p*’ has as its meaning the fact *aRb*,” we will have: “‘*p*’ says that *aRb*, and *aRb*.” In “Facts and Propositions” Ramsey indicates how these first steps are to be developed. For our immediate purposes, though, the first steps are enough. They already show that, in Wittgenstein’s view, *neither* a proposition’s having a certain content (its possession of a truth condition) *nor* its being true (its truth condition’s being satisfied) is to consist in the proposition’s connecting with any kind of “thing.”

This confirms our negative contention in this section: Wittgenstein’s theorizing, from 1913 onward, simply makes no room for a correspondence theory of truth. But now, what *positive* view of truth do Wittgenstein’s considerations suggest? To answer this question it is useful to return to the very beginning of the story we have outlined. In Russell’s early theory both the “sense of a proposition” (what Russell then called an abstract “proposition”) and the “meaning of a proposition” (what Russell then called a “fact”) were things: in the case of a true proposition they were, indeed, the *same* thing. Russell’s later theory then introduced an asymmetry between his accounts of what it is for a proposition to have a certain content and of what it is for the proposition to be true: the second is, but the first is not, a matter of its connecting with a unitary complex entity (a “thing”). It is by introducing this asymmetry that Russell repudiates his early “identity theory” of truth. Wittgenstein, as we have just seen, undoes this step, and so restores symmetry between the accounts of content and truth. He thereby opens up a new position in the space of theoretical alternatives, one that accommodates a new form of the identity theory.

To see that this is more than an abstractly characterized possibility, we need only reflect on the motivations of the identity theory. Moore wrote in 1902 that what refutes the correspondence theory is “the impossibility of finding any difference between a truth and the reality to which it is supposed to correspond” (1902: 717). Russell echoed this in 1904, holding that a true (abstract) proposition cannot be distinguished from the supposedly corresponding fact (1904: 473). Wittgenstein echoed it again in the *Tractatus*, holding that “What the picture represents is its sense” (TLP 2.221), so that “Instead of, This proposition has such and such a sense, we may say, This proposition represents such and such a situation” (4.031). What all these formulations share is an insistence that there can be no distinction between what a proposition affirms and what obtains if the proposition is true. Where they differ is only in the grammatical construction they place

on this claim. Moore and Russell construe it as an identity statement, understanding such phrases as “what a proposition affirms” as singular terms. Wittgenstein, by contrast, holds that such phrases are “incomplete symbols” or, in the more familiar terminology of ordinary grammar, that they formulate indirect questions. Thus a Wittgensteinian “identity theory” asserts an identity, not between two species of abstract entity, but between two sorts of question. To ask after the content of a proposition, and to ask what is required for its truth, are, according to Wittgenstein’s identity theory, two ways of asking the same thing. If we have an answer to the first of these questions, then there is nothing that an answer to the second need, or even can, add to it.

6.4 RAMSEY

6.4.1 Overview

The conclusion just reached, though we presented it as representing Wittgenstein’s view, also represents the right understanding of the most often quoted passage of Ramsey’s “Facts and Propositions,” in which he asserts that “there is really no separate problem of truth” (1927: 142).¹⁷ “No *separate* problem”—we emphasize—taking him to mean that there is no problem about what it is for a judgment (or proposition) to be true that can be separated from the question of what it is for the judgment to have a certain content.

On Ramsey’s reasons for this view we can be brief, since in essentials they recapitulate the Wittgensteinian considerations already reviewed in section 6.3. In section 6.4.2 we will explain how in the opening section of “Facts and Propositions” (138–42) Ramsey deepens and extends Russell’s argument for a multiple-relation theory, and will review the new justification he offers for Wittgenstein’s contention, that phrases apparently referring to Russell’s “facts,” as well as his early “propositions,” should be construed as “incomplete symbols.” Then, in section 6.4.3, we will outline, and begin to counter, the very widely held view that Ramsey advanced a “redundancy theory” of truth. Our consideration of this view will, however, not be complete until the final section of the chapter, where we comment on what is probably the strongest reason for interpreting Ramsey as a redundancy theorist, namely, the role that this interpretation suggests for the “pragmatist” elements that Ramsey introduces into the theory of content. In his essay “Truth” (1959) Michael Dummett famously and influentially argued that a redundancy theory of truth cannot be combined with a truth-conditional account of content, one that accords with Wittgenstein’s statement, quoted above, that “to understand a proposition is to know what is the case if it is true” (1913: 104). Now suppose—as is very plausible—that Ramsey already appreciated Dummett’s point. Then, if Ramsey had endorsed the

¹⁷ Further references to “Facts and Propositions” in section 6.4 will be given by page number alone.

redundancy theory, he would need an alternative, non-Wittgensteinian account of content, one that explains what it is for a judgment or proposition to have a certain content independently of any appeal to the notion of truth. According to the interpretation we are considering, Ramsey's pragmatism supplies precisely this need. A proper assessment of this interpretation would require a detailed treatment of the pragmatist theory of content it involves, and hence would carry us too far from the advertised theme of the chapter. Instead we will have to content ourselves in section 6.4.4 with indicating how Ramsey's pragmatist innovations might be understood as complementing, rather than replacing, a basically Wittgensteinian approach to the theory of content.

The interpretation just mentioned emphasizes the need to account for how Ramsey conceived the relations amongst the various proposals advanced in "Facts and Propositions." It is right to do so. But we think the best way of meeting this need is to account for how Ramsey understood his proposals as relating to the theories of his predecessors. His essay opens with a question about the "logical analysis" of belief, posed very much in the manner of Russell, who had spoken of "the problem of the logical form of belief, i.e. what is the schema representing what occurs when a man believes" (1922: 19). It ends with the very generous acknowledgment to Wittgenstein that we mentioned in our Introduction: "Everything that I have said is due to him, except the parts that have a pragmatist tendency, which seem to me to be needed to fill up a gap in his system" (155). Taking these facts as a guide to Ramsey's intentions, what we should expect to find in the essay is a largely Wittgensteinian answer to a Russellian problem. In our view, this is just what the essay contains.

6.4.2 Ramsey's multiple-relation theory

We observed in section 6.3.4 that Russell's thesis, that propositions are incomplete symbols, is primarily a reductive ontological claim. In this respect the thesis represents a departure from Russell's earlier applications of the notion of an incomplete symbol, canonically in his theory of descriptions and in the treatment of class terms offered by his "no classes" theory (Whitehead and Russell 1910: 66–84). In those earlier applications the essence of the claim that phrases of a certain sort are incomplete symbols is that those phrases "never have any meaning in themselves" (Russell 1905a: 416), or again, that such a phrase "is not supposed to have any meaning in isolation, but is defined only in certain contexts" (Whitehead and Russell 1910: 67). The significance of the phrase will therefore not be specified by identifying what (the thing which) it means. Instead, its significance will be determined by a contextual definition, a rule for paraphrasing any sentence in which the phrase occurs into a sentence of a different form which will include no unitary constituent corresponding to the original phrase. Plainly, this understanding of the notion cannot apply to all of those phrases that Russell had formerly taken to have abstract propositions as their meanings. Those phrases are, or at least include, *sentences*, and it would be patently silly to suggest that sentences should be dissolved by eliminative paraphrase into their sentential contexts.

Russell of course does not advance that absurd suggestion. He nonetheless attempts to minimize the change to his earlier conception of an incomplete symbol that is forced by this new application of the notion. The attempt is unpersuasive:

The phrase which expresses a proposition is what we can call an “incomplete” symbol; it does not have meaning in itself, but requires some supplementation in order to acquire a complete meaning. This fact is somewhat obscured by the circumstance that judgement in itself supplies a sufficient supplement, and that judgement in itself makes no *verbal* addition to the proposition. Thus “the proposition ‘Socrates is human’” uses “Socrates is human” in a way which requires a supplement of some kind before it acquires a complete meaning; but when I judge “Socrates is human”, the meaning is completed by the act of judging, and we no longer have an incomplete symbol. (Whitehead and Russell 1910: 44)

According to his new, multiple-relation theory, when *s* judges that Socrates is human, a fact exists in which *s*, Socrates, and humanity are combined by the relation of judgment. It is this fact which Russell proposes as the “complete meaning” acquired by “Socrates is human” when it is supplemented by *s*’s act of judgment. But this fact cannot be a “complete meaning,” since it is not a meaning at all. Specifically, it cannot be the meaning of “Socrates is human” when this is judged by *s* (since that *s* so judges is no part of what *s* judges to be so), and it cannot be the meaning of “*s* judges that Socrates is human” (since this sentence would then be meaningless unless true).¹⁸ Of course, what Russell *ought* to say is, simply, that his new theory requires no such (unitary) things as “the meanings” of sentences. Instead he tries to find, from amongst the entities his theory of complexes admits, replacements for them. This is an indication of the primarily ontological, rather than analytical character of Russell’s new theory. It is a consequence of the way this theory is centrally motivated, by the ontological problem of false propositions. And it is the feature of Russell’s theory that Ramsey targets in the opening arguments of “Facts and Propositions.”

We noted in section 6.3.3 that, while the problem of false propositions led Russell to reject a dyadic conception of judgment as “impossible,” he retained a dyadic conception of perception: while *s*’s judgment that the knife is to the left of the book will be represented by the schema, $J(s, k, L, b)$, the perception on which this judgment is based remains of the form, $P(s, <kLb>)$. Ramsey’s first substantial move¹⁹ is to argue

¹⁸ Might the fact that Russell has in mind—or at least something very like it—be the meaning of (the thing meant by) another kind of phrase, for instance, a singular term such as “*s*’s judgment”? Yes. We will see later in this section that Ramsey admits this suggestion; we will also see that it reinforces, rather than undermines, the point currently at issue.

¹⁹ Before this Ramsey allows that the narrow problem of false propositions might be avoided by a theory according to which each of a pair of contradictory judgments has reference to the same fact—one of them, as it were, “pointing towards” this fact, the other “pointing away” from it (139–40, referring to Russell 1921: 272). We saw in section 6.3.4 that this is how Russell understood Wittgenstein’s 1913 view, and also that it represents a misunderstanding of that view. Ramsey’s concession to this theory forms a minor part of his case that the specific problem of *false* propositions is not the proper ground for the multiple-relation theory.

that a dyadic analysis “*either* of judgement *or* of perception” is inadequate (140; emphasis added), and for the *same* reason. If successful, this first argument will show that, in motivating his multiple-relation theory by the problem of false propositions, Russell misidentified the issue. That perception is factive but judgment not—or in the material mode, that a perception is “infallible” while a judgment may be false—is not the relevant point.

The argument begins by posing the question, whether “He perceives that the knife is to the left of the book” asserts a dyadic relation between a person and a fact. It is short, and decisive:

Suppose that I who make the assertion cannot myself see the knife and book, that the knife is really to the right of the book, but that through some mistake I suppose that it is on the left and that he perceives it to be on the left, so that I assert falsely “He perceives that the knife is to the left of the book.” Then my statement, though false, is significant, and has the same meaning as it would have if it were true; this meaning cannot therefore be that there is a dual relation between the person and something (a fact) of which “that the knife is to the left of the book” is a name, because there is no such thing. (140)

In brief, and accepting that perception is “infallible,” the *truth* of “*s* perceives that *p*” entails the truth of “*p*”; but it is not the case, as Russell’s dyadic analysis would imply, that the *significance* of “*s* perceives that *p*” entails the truth of “*p*.”

This argument should be compared to the criticism, given independently above, of Russell’s suggestion in *Principia* that the fact of a subject’s making a certain judgment might serve as a “complete meaning.” This fact is the only candidate Russell’s theory of complexes provides to be the meaning of a statement that the subject so judges. Similarly, the fact of perception would be the only candidate to be the meaning of a statement that the subject enjoys a certain perception. In each case, though, the effect of electing this sole candidate would be that the statement becomes meaningless unless true. Perception is factive, judgment not. But *that* someone enjoys a certain perception, just as much as that he makes a certain judgment, is something that can be claimed to be so when it is not. Thus, just as what is claimed in the second case cannot be identified with any entity supplied by the theory of complexes, so likewise what is claimed in the first case cannot be constructed from materials supplied by that theory.

In this first argument we already see that Ramsey’s approach to the issue contrasts with Russell’s in just the same way that Wittgenstein’s approach also contrasts with Russell’s (see again section 6.3.4). Russell in effect asks, “What kind of occurrence is a perception, or a judgment?” For answering a question posed in this ontological mode it was surely reasonable to think, as Russell did, that the difference between perception, as factive, and judgment, as non-factive, would be significant. Ramsey instead asks, in analytical mode, “How should we understand the use of phrases serving to specify the content of a perception or judgment?” For *this* question, as Ramsey’s argument shows, the difference between factive and non-factive attitudes is irrelevant.

This contrast in approach stands out still more clearly, though, in Ramsey's second argument—where, we should note, Ramsey is clearly no longer borrowing from Wittgenstein but moving beyond him. The conclusion from the first argument is that “[a phrase such as] ‘that the knife is to the left of the book’, whether it is true or false, cannot be the name of a fact” (140). Ramsey now turns to the alternative suggestion that a phrase of this kind functions as a definite description of a fact. According to this suggestion, a sentence containing the phrase would be *significant* even if the phrase were empty, i.e. even if there were no fact meeting the description it expresses. We would then no longer face the consequence that the significance of “*s* perceives that *p*” entails “*p*.” Instead, a phrase “that *p*” would be exactly comparable to a description of an event, such as “the death of Caesar”: someone might significantly (though of course falsely) have said, “He witnessed the death of Caesar,” while Caesar yet lived.

What for our purposes is most important about Ramsey's response to this suggestion is that he does not enter into any dispute as to whether there are entities (“events”) for descriptive phrases of this kind to denote. Nor does he dispute that phrases of this kind are often reasonably construed as denoting such entities. Instead, Ramsey draws attention to the contrasting logical behavior of *different uses* of such phrases. A context in which “the death of Caesar” functions as a definite description (and where no further operators create ambiguities of scope) is extensional: if *s* witnessed, or was present at, or caused, or took part in the death of Caesar, then *s* witnessed, or was present at, or caused, or took part in the murder of Caesar, since those were one and the same event. But when the same phrase is used in “the discussion of cognition”—when, for instance, “He was aware of the death of Caesar” is used with the meaning of “He was aware that Caesar had died”—the context is non-extensional: “he could quite well be aware that Caesar had died without knowing that he had been murdered” (141).²⁰ Just as we saw with Wittgenstein, then, Ramsey's focus in this argument is not on a range of putative entities, but on the logical functioning of phrases that Russell had construed as denoting such entities. Independently of any ontological contention as to whether there are things—be they “events,” “complexes,” “facts”—for these phrases to refer to, Ramsey's conclusion is that, in its use in specifying the content of cognition, “a phrase beginning ‘the fact that’ is not a name, and also not a description,” of any such thing (141). With this argument, then, Ramsey finally separates the multiple-relation theory from the style of ontological theorizing that originally motivated Russell to adopt it. Furthermore, while parts of the argument are clearly original with Ramsey, the conclusion it yields—that “a proposition about ‘the fact that *aRb*’ must be analysed into (1) the proposition *aRb*, (2) some further proposition about *a*, *R*, *b*, and other things” (141–2)—is exactly the view that Ramsey had earlier ascribed to Wittgenstein (1923: 273).

²⁰ Is “He perceived the death of Caesar” an extensional context for “the death of Caesar”? Perhaps so. If so, it is not equivalent to “He perceived *that* Caesar had died.”

6.4.3 Is Ramsey a redundancy theorist?

The first claim of a redundancy theory is that explicit predication of truth is redundant—that nothing can be said with it that is not at least as well said without it. This much Ramsey clearly holds. The sentence, “It is true that Caesar was murdered,” he says, “means no more than that Caesar was murdered” (142). Some disquotationalists (e.g. Quine 1970) have held that a truth-predicate is eliminable in this kind of way only if the proposition to which it is ascribed is spelled out, and that the usefulness of the predicate shows itself when it is ascribed to propositions that are merely described or spoken of generally, as in “Everything he says is true.” Ramsey goes further, holding that “true” is eliminable even from these contexts. In “For all p , if he asserts p , p is true,” he says, the need for the verb-phrase “is true” is only an imposition of English grammar, which does not cope readily with generalization into sentence positions (143).

As Wittgenstein remarked, though, to do away with the words “true” and “false” is not to “do away with the puzzles connected with truth and falsity” (1979b: 106). The redundancy theorist is distinguished, not merely by advancing the eliminability claim, but by the explanation he offers for it: he holds that the word “true” is eliminable without loss because there is no substantial notion for it to express. An alternative suggestion, due to Frege (1984: 354), is that explicit predication of “true” *adds* nothing because the notion it expresses is *already present* in that to which it is ascribed.

Before we count Ramsey a redundancy theorist, then, we need to attend to more than the two paragraphs (at 142–3) in which he presents the eliminability claim. In the previous section we considered the engagement with Russell that leads up to this claim; we saw that Ramsey’s aim, like Wittgenstein’s, was to separate the grounds of the multiple-relation theory from any involvement with an ontology of facts. In the remainder of this one we assemble some observations about the immediate use to which Ramsey puts this claim.

Here the most immediately telling fact is one we have already noted. Whereas a redundancy theorist might claim that there is really no problem about truth, Ramsey instead holds that “there is really no *separate* problem of truth” (142; emphasis added). In Ramsey’s argument the immediate role of the eliminability claim is to redirect misguided enquiry as to “the nature of truth” onto its proper object, “the nature of judgement or assertion.” It serves, for instance, to convert “His judgment is true” into “If he asserts p , then p ,” and plainly “what is difficult to analyse in [this] formulation is ‘he asserts p ’” (143; variable altered).

A second telling fact is that in reformulating this point Ramsey shows no sign of shying away from, but instead seems clearly to endorse, the Tractarian view that for a judgment to have a certain content is for it to have a certain truth condition:

It is, perhaps, also immediately obvious that if we have analysed judgement we have solved the problem of truth; for . . . the truth or falsity of [a judgment] depends only on what proposition it is that is judged, and what we have to explain is the meaning of saying that the judgement is a judgement that a has R to b , *i.e.* is true if aRb , false if not. (143)

This should not be surprising, since in this paragraph Ramsey is simply repeating what he had said four years earlier, in exposition of Wittgenstein, in his Critical Notice of the *Tractatus*:

First, it may be remarked that if we can answer our question [the question Ramsey has just formulated, “What is it for a proposition token to have a certain sense?”] we incidentally solve the problem of truth; or rather, it is already evident that there is no such problem. For if a thought or proposition token “ p ” says that p , then it is called true if p , and false if $\sim p$. (1923: 275)

It is relevant to observe that in his lectures, too, Ramsey repeatedly presented instances of the eliminability claim as illustrations of *Wittgenstein’s* analyses.

A third point worth noting is that Ramsey’s eliminative paraphrases make use of a variable “ p ” in sentence position (one “ranging over propositions”). Now Ramsey would be the last person simply to help himself to a variable without considering the question, what determines its range, or what notions are involved in grasping the generality it expresses. The subsequent argument of “Facts and Propositions” is structured by Ramsey’s Tractarian answer to this question: the variable ranges over all truth-functions of elementary propositions.

Finally we should note that, at the conclusion of this short passage on the problem of truth, Ramsey concentrates the critical force of his reflections against any explanatory appeal to the notion of a “fact”:

... what we have to explain is the meaning of saying that the judgement is a judgement that a has R to b , *i.e.* is true if aRb , false if not. We can say, if we like, that it is true if there exists a corresponding fact that a has R to b ; but this is essentially not an analysis, but a periphrasis, for “the fact that a has R to b exists” is not different from “ a has R to b .” (143)

The most obvious point here is that Ramsey follows Wittgenstein (as we interpreted him in section 6.3.3) in denying any theoretical weight to talk of the existence of facts. Such talk is, he very clearly says, no more than long-windedness. He thereby rules out any appeal to the notion in the service of a correspondence theory of truth. What perhaps stands out less obviously, but is more important for the overall story we have been telling, is the context in which Ramsey envisages this appeal being made. Talk of the existence of a corresponding fact is, he suggests, one of the things we might uselessly resort to through failure to answer the real question: what it is for a judgment to have the content that aRb . In section 6.2.3 we suggested that Russell makes his resort to a correspondence theory of truth to compensate for the inadequacy in his new theory of judgment. We were at that point following exactly in Ramsey’s footsteps. Ramsey’s view is not only that the correspondence theory is a bad answer, but that it is a bad answer to the wrong question.

These observations support the conclusion that, in the paragraphs of “Facts and Propositions” that have so often been cited as advancing a redundancy theory, Ramsey

took himself to be developing views already advanced by Wittgenstein; and, on the account of Wittgenstein's views offered in section 6.3, Ramsey was clearly right in this supposition. There is, then, no better reason to attribute a redundancy theory to Ramsey than to Wittgenstein.

6.4.4 Ramsey's pragmatism

Wittgenstein wrote in the *Tractatus*:

The proposition is the expression of its truth-conditions.

(Frege has therefore quite rightly put them at the beginning, as explaining the signs of his logical symbolism . . .)²¹ (TLP 4.431)

We have so far seen no reason to distance Ramsey from this Tractarian conception. Putting together the results of the previous two sections we can say that Ramsey's aim, like Wittgenstein's, was to refuse any explanatory role, in the analysis of cognition and hence in the analysis of truth, to the notion of a "fact." This is a repudiation of the correspondence theory of truth, not of the notion of truth itself, nor of its place "at the beginning" of a theory of content. It would be entirely reasonable to complain, though, that we have so far considered only a part of Ramsey's essay. In this final section we offer a response to this complaint.

Adapting Ramsey's eliminative paraphrases one very soon arrives at the following principles governing ascription of truth and falsity to a belief:

B is true iff $\exists p$ (B is a belief that p , and p);

B is false iff $\exists p$ (B is a belief that p , and not p).

Suppose that these principles were advanced, in the spirit of the redundancy theory or similar deflationary views, as explaining all that needs to be explained as to what it is for a belief to be true. Since these principles presuppose a prior understanding of a belief's being a belief that p , we could not then explain *this* notion—the notion of a belief's having a particular content—by reference to the condition for the belief to be true: we would need an alternative, non-truth-conditional theory of content. According to the style of interpretation mentioned in section 6.4.1 this theoretical need is met by Ramsey's pragmatist proposal that "the meaning of a sentence is to be defined by reference to the actions to which asserting it would lead, or, more vaguely still, by its possible causes and effects" (155). Moreover, the very fact that he offers this pragmatist theory is

²¹ Omitted here is a criticism, one of several Wittgenstein makes, of the specific form of Frege's explanations. Right or wrong, these specific criticisms are of much less importance than the agreement against which they are set.

the strongest reason for attributing to Ramsey the commitments about truth that create the need for it. The redundancy theory of truth (or perhaps a less stringent deflationism) forms, along with a pragmatist theory of content, a coherent theoretical package. Unless we attribute to Ramsey the first element of this package, the thought goes, we will be at a loss to explain what the second element is *for*.

Much of what we have to say against this style of interpretation has already been said in the two previous sections. Ramsey explicitly remarks that his pragmatist “additions” represent a departure from Wittgenstein (155). The interpretation offers to make sense of the overall structure of “Facts and Propositions” by representing this acknowledged departure as necessitated by an earlier and more fundamental departure from Wittgenstein regarding the notion of truth. The suggestion thus falsifies Ramsey’s clear statement that his pragmatism is the *only* non-Wittgensteinian element of the essay. More importantly, it makes a mystery of the fact, evidenced above, that in his discussions of judgment and truth Ramsey is advancing views that he elsewhere attributes to Wittgenstein. Insofar as the interpretation is recommended as making the best overall sense of Ramsey’s intentions, these facts significantly undermine its appeal.

Ramsey’s “very vague and undeveloped” (155) pragmatism has inspired the approach to the theory of content known as “success semantics,” most fully articulated by Mellor (2012), and the “exclusionary semantics” of Rumfit (2011). It is not possible, and anyway it would not be relevant, to review here the strengths of these theories in themselves. Nor should our question be whether, in their developed forms, these theories can claim to represent Ramsey’s intentions: both authors are completely clear that they are developing Ramsey’s ideas in ways that he did not anticipate in any detail. It *is* relevant to ask, though, whether the general shape of these theories suits them to “fill up the gap in [Wittgenstein’s] system” that Ramsey intended his pragmatism to fill.

We need first a description of the gap. Ramsey provides this in his Critical Notice, where the course of discussion anticipates and closely mirrors that in “Facts and Propositions.” As there, Ramsey holds that the “problem of truth” will be solved by a correct theory of the content of judgment; and he endorses the Wittgensteinian form of the multiple-relation theory—according to which “A believes *p*” has the form of “*p*’ says that *p*” (TLP 5.542)—as having reduced the latter problem to the question, “What is it for a proposition token [a sentence] to have a certain sense?” (1923: 275). The picture theory gives Wittgenstein’s answer to this question:

the sense of a proposition is that the things meant by its elements (the words) are combined with one another in the same way as are the elements themselves, that is, logically. (ibid.)

This answer Ramsey judges to be seriously “incomplete,” in that it is applicable only to “completely analysed elementary proposition[s]” (ibid.) He passes quickly over the issue of how it is to be adapted to the unanalyzed propositions of colloquial language, anticipating (rightly) that Wittgenstein would dismiss the intricacies of ordinary speech as of no more than empirical interest. But he insists that a second issue, of how the picture

theory extends to non-elementary propositions, “must be faced” (1923: 276). Ramsey then provides a succinct account of Wittgenstein’s theory of “the sense of propositions in general” as truth-functions of elementary propositions, spelled out first for finite truth-functions and then for quantified propositions. But this theory, he holds, is certainly not an adequate response to the question posed,

For it is really only an account of what senses there are, not of *what propositional signs have what senses*. (1923: 277; emphasis added)

This, clearly, is “the gap.” Wittgenstein holds that non-elementary propositions express truth-functions of elementary propositions, but he offers no account of *how* they do so. Obviously, the expression of non-elementary senses is achieved by the use of “logical constants” (Ramsey here means the *words* “not,” “or,” “all,” etc.): somehow the constants serve to “complete the determination of the sense,” but the way in which they do this “is left obscure” (1923: 288). How, then, do these words work?

The Critical Notice does not reach an answer to this question, but from the kinds of approach Ramsey canvasses we can draw further confirmation that this is the question that the pragmatist elements of “Facts and Propositions” are designed to answer. He first considers and rejects the Wittgensteinian suggestion that the various propositional signs expressing a single sense should be those constructed according to a particular syntactic rule: even if such rules were possible,²² the approach could only yield language-relative answers to a language-neutral problem. A Russellian suggestion, connecting the constants to “special belief feelings,” would avoid this language-relativity: “Logical constants might then be significant as substitutes for these feelings, which would form the basis of a universal logical symbolism of human thought” (1923: 278); this approach, though, would need to be complemented by an account of equivalence amongst different configurations of these feelings. The 1923 discussion breaks off here, with Ramsey merely pointing to some of the obstacles in the way of such an account (1923: 279). “Facts and Propositions” resumes the discussion. It again suggests that the constants have meaning as substitutes for special feelings: to take one central instance of this proposal, the word “not” will be viewed as having acquired meaning through “a sort of association” that renders it “part of the internal language of a speaker,” whereby feeling *disbelief* toward “*p*” and feeling belief toward “not *p*” come to be interchangeable, equivalent occurrences (148). And it adds the distinctively pragmatist contention that “[this] equivalence . . . is to be defined in terms of causation, the two occurrences having in common many of their causes and effects” (149).

On this account of “the gap,” does either of the pragmatist theories mentioned have the right general shape to fill it? In the case of success semantics, it seems clear that it

²² Ramsey rightly doubts this, saying that such a rule “seems to presuppose the whole of symbolic logic” (1923: 278): since “*p*” is equivalent to “*p* and Tautology,” a rule determining *all* the equivalents of “*p*” would be a decision procedure for logical truth.

does not. This approach draws on Ramsey's famous example of the chicken, about which we might say that "it believes a certain sort of caterpillar to be poisonous, and mean by that merely that it abstains from eating such caterpillars on account of unpleasant experiences connected with them" (144). Generalizing the example:

any set of actions for whose utility p is a necessary and sufficient condition might be called a belief that p , and so would be true if p , i.e. if they are useful. (ibid.)

The example thus leads directly to the behaviourist "chicken equation":

to believe that $p =$ to ϕ iff (it is useful to ϕ iff p).

Success semantics modifies this, identifying a belief not with a pattern of behavior, but with the mental state that causes it. It thus proposes to identify the content of a belief with the circumstance in which actions which would be caused by the belief, in conjunction with various desires, would be successful in achieving the aims of those desires.

When we consider this theory as a possible realization of Ramsey's intentions, the most obvious point to make is that the example it builds on is offered by him as illustrating a kind of belief he does *not* intend to discuss (144). Perhaps not too much should be made of this point on its own: although Ramsey sets aside "chicken beliefs" at the start, there are passages later in the essay that clearly draw on the "chicken equation"—as, for instance, when he raises the question of the value of true beliefs (148).²³ But much greater weight attaches to Ramsey's elaboration of the point:

I prefer to deal with those beliefs which are expressed in words, or possibly images or other symbols, consciously asserted or denied; for these beliefs, in my view, are the most proper subject for logical criticism. (144)

Beliefs expressed in words have their content in virtue of the compositional structure of the sentences used to express them. In the case of an atomic sentence, " aRb ," Ramsey clearly sets compositionality as a constraint on any account of the causal powers of a belief expressed by it.²⁴ And the central task of the paper is then to characterize on that basis the causal properties of beliefs expressed by logically complex sentences. In fundamental contrast to this, success semantics specifies a condition for something to have a certain content that is independent of attributing any compositional structure to it.

²³ The "chicken equation" directly yields "It is useful to believe that p iff p ." (Substitute its right-hand side for its left in the context "It is useful . . .," and observe that the result, being of the form " q iff (q iff p)," is tautologically equivalent to " p "; see 144 fn. 1.)

²⁴ Ramsey stipulates (149) that these causal properties must be "connected with a , R , and b in such a way that the only things which can have them must be composed of names for a , R , and b ." (This is the doctrine that the meaning of a sentence must result from the meaning of the words in it.)

This is not of course to say that the approach cannot be developed to account for the content of linguistically structured representations. It does imply, though, that in its basic conception success semantics is not designed to address the specific problem for which Ramsey sought a pragmatist solution.

Things stand differently with the “exclusionary semantics” proposed by Rumfit, an explicitly compositional theory. This theory draws its inspiration from a later stage in the argument of “Facts and Propositions,” where Ramsey is discussing the attitudes expressed by truth-functionally complex sentences:

To say that feeling belief towards a sentence expresses such an attitude is to say that it has certain causal properties which vary with the attitude, i.e. with which possibilities are knocked out and which, so to speak, are still left in. Very roughly the thinker will act in disregard of the possibilities rejected, but how to explain this accurately I do not know. (150)

In context, “possibilities” are the truth-possibilities of Tractarian elementary propositions. An “attitude” is specified as being in agreement with some set of these possibilities and in disagreement with the rest; an attitude, that is, is what Wittgenstein called a *sense*. Ramsey’s proposal is to connect with each such attitude those propositional signs (sentences) whose acceptance has causal properties appropriate to the attitude, thus answering the question “what propositional signs have what sense” (1923: 277).

Rumfit however abstracts from this Wittgensteinian context to distill the more general idea that “a statement’s content is determined by the possibilities it excludes,” meaning those which someone making or accepting the statement is committed to exclude (2011: 230). Possibilities are understood to be “extra-linguistic things” specified by clauses (233); and to exclude any possibility is thereby to exclude any more determinate possibility which would, of necessity, realize it (234). (For instance, to accept “The plate is round” is to exclude the possibility that the plate is square, and thereby to exclude its being nine-inches square.) With these framework ideas in place Rumfit develops in the central section of his paper an exclusionary semantics for the language of propositional logic. The semantics associates with each atomic statement the set of possibilities it excludes.²⁵ Non-classical clauses for the connectives then represent the possibilities excluded by any compound statement as a function of those excluded by its immediate constituents.

This minimal sketch—inadequate as it would be for an assessment of Rumfit’s theory in itself—allows us to ask whether a theory of this broad shape is suited to address the question Ramsey posed. From comparison with the passage that inspires it (quoted earlier) it seems plain that it is not. The proposal offers an alternative construal of the possibilities excluded by basic statements, and thus a different construction of the “attitudes”

²⁵ The paper includes no discussion of how the semantic values of atomic statements (the sets of possibilities excluded by them) are determined by those of their constituent words. Hence our comments will likewise abstract from Ramsey’s first compositionality constraint.

definable by reference to them. But when we reach the notion Ramsey confessed himself unable to explain—what it amounts to, in causal terms, to act in disregard of excluded possibilities—we are, surely, no further forward. In effect, then, exclusionary semantics offers a replacement for the part of the job which, in Ramsey's view, Wittgenstein had already done, and very little to make good what Wittgenstein had left undone. It seems likely that Ramsey would have said of this approach, as he said of Wittgenstein's, that it is "really only an account of what senses there are, not of what propositional signs have what senses" (1923: 277).

Modulo an account of the contents of atomic sentences, a theory of this kind supplies, for each sentence of the language with which it deals, a specification of what content it has. If this were enough to fill "the gap," though, there could have been no gap; for Ramsey of course had available, in the usual classical clauses for the logical constants, something to do exactly this job. The gap Ramsey identified needs to be filled, not with a specification of what content a given sentence has, nor yet with a statement of what in general the content of a sentence consists in (of "what senses there are"), but with an account of what features those are in virtue of which any particular sentence expresses such a content (of "what propositional signs have what sense"). It helps to clarify his understanding of this requirement to note that, when compared to the issue addressed by a systematic, compositional semantics, Ramsey's question is posed, as it were, in reverse. Instead of demanding, for each sentence, a specification of its content, Ramsey is seeking, for each content, a specification of those sentences which have it. The guiding thought behind this formulation is that we will understand what it is for any sentence to have such a content just in case we can characterize the feature that is shared in common between sentences having the same content. We saw that the Critical Notice rejects a syntactical approach to this question, attributed to Wittgenstein (1923: 278). Early in the discussion of "Facts and Propositions" Ramsey likewise rejects an ontological approach to it, attributed to Chadwick, according to which equivalent complex sentences would present, as it were, the same logical configuration of elementary states of affairs (1927: 146). The decisive step in Ramsey's pragmatism is to turn, with Russell, to psychology for an answer to this question, locating the commonality between sentences sharing a sense in the causal role their acceptance plays in a subject's mental economy in general, and more specifically in the guidance of his actions.

To take this step is not to turn away from Wittgenstein. After all, when Russell put to Wittgenstein a similar though less general question, about the constituents of a "thought" and their reference—or, in Ramsey's way of speaking, about the composition and content of "non-verbal tokens" of a proposition (1923: 274)—he replied that "it would be a matter of psychology to find out" (Wittgenstein 2008: 99). There is of course an important difference: referring it to psychology was for Wittgenstein a way of dismissing the problem, while for Ramsey it is a proposal for solving it. But it is one thing to acknowledge that Ramsey is taking a step beyond anything attempted by

Wittgenstein, quite another to suggest that this step contradicts what Wittgenstein had already done.

Our consideration of Ramsey's pragmatism in this section—like the accounts of his discussions of judgment and truth given in the previous two—has turned up nothing to support this suggestion. Our discussion has been highly selective: we have not engaged at all with the many obstacles Ramsey identified to the detailed working out of a pragmatist theory, but have aimed only to characterize the role his pragmatism plays in the overall structure of the argument of "Facts and Propositions." For the purpose in hand, though, this selective treatment is enough. It is on just such broad, structural grounds that the style of interpretation outlined at the start of this section presents Ramsey's advocacy of pragmatism as a reason to attribute to him a redundancy theory of truth. Thus to show that such an interpretation misidentifies the gap that pragmatism is to fill is to undercut that reason. And so our previous conclusion can stand: there is no good reason to attribute to Ramsey a redundancy theory of truth.

6.5 CONCLUSION

The story we set out to tell, spanning as it does some twenty-five years of profound thought expressed in some far from easy texts, has been rather a long one. So it might be useful to gather together at the end a stark summary of our conclusions.

Regarding Russell we have claimed:

- (i) Russell's rejection of his early theories, of judgment as a dyadic relation between a subject and an objective proposition, and of truth as a simple property of propositions, was centrally motivated by the ontological problem of false propositions.
- (ii) There is no essential or necessary connection between the two components of the position Russell then adopted, the multiple-relation theory of judgment and the correspondence theory of truth.
- (iii) Instead, these components are connected in Russell primarily through the style of ontological theorizing in which they are developed. His general theory of complexes supplies no configuration that might possess intrinsically, in virtue of its own internal form, the characteristic features of judgment—of having content and being true or false. Hence Russell aims to explain these features relationally, through a correspondence between one complex, the truth-bearing judgment, and another, the truth-making fact.

Regarding Wittgenstein we have claimed:

- (i) Wittgenstein saw (a) that the two components of Russell's new position are separable, (b) that the first component, Russell's version of the multiple-relation theory, is inadequate as an explanation of what it is for a judgment to have content, (c) that Russell appealed to the second component, his correspondence theory of truth, in an attempt to make good that inadequacy, and (d) that rectifying the inadequacy of the first component would render the second component unnecessary.
- (ii) The picture theory of the *Tractatus* is Wittgenstein's revised version of the multiple-relation theory.
- (iii) The picture theory leaves no room for a correspondence theory of truth.

Regarding Ramsey we have claimed:

- (i) Ramsey's remarks on judgment and truth are presented in criticism of a correspondence theory of truth, as advanced by Russell from 1910.
- (ii) In presenting this criticism, Ramsey supposed himself to be developing views already advanced by Wittgenstein, and in this supposition he was entirely right.
- (iii) The extent of agreement between Wittgenstein and Ramsey counts definitively against attributing to Ramsey a redundancy theory of truth.
- (iv) The pragmatist theory of content that Ramsey introduced in "Facts and Propositions" provides no good reason to revise that judgment. Ramsey's pragmatism is an addition to Wittgenstein's approach, not a replacement for it: it is not an alternative to the truth-conditional conception of what the content of a judgment or sentence consists in, but a naturalistic account of the features that confer on a judgment or sentence such a content.

Our aims have been exegetical. We have tried to describe how Wittgenstein and Ramsey rejected the correspondence theory of truth. We have not tried to show that the correspondence theory *should* be rejected. It will no doubt have been obvious, though, that that is what we think.

REFERENCES

- Bostock, D. (2012). *Russell's Logical Atomism*. Oxford: Oxford University Press.
- Cartwright, R. (1987). "A Neglected Theory of Truth," in R. Cartwright, *Philosophical Essays*. Cambridge, MA: MIT Press, 71–93.
- Dummett, M. (1959), "Truth," reprinted in his *Truth and Other Enigmas* (1978). London: Duckworth.
- Frege, G. (1979). *Posthumous Writings*, trans. P. Long and R. White. Oxford: Blackwell.
- Frege, G. (1984), *Collected Papers on Mathematics, Logic, and Philosophy*, ed. B. F. McGuinness. Oxford: Blackwell.

- Hanks, P. W. (2012), "Early Wittgenstein on Judgement," in J. L. Zalabardo (ed.), *Wittgenstein's Early Philosophy*. Oxford: Oxford University Press, 37–63.
- Mellor, D. H. (2012). "Successful Semantics," in D. H. Mellor, *Mind, Meaning, and Reality*. Oxford: Oxford University Press, 60–77.
- Moore, G. E. (1899). "The Nature of Judgement," reprinted in T. Regan (ed.), *G. E. Moore: The Early Essays* (1986). Philadelphia: Temple University Press, 59–60.
- Moore, G. E. (1902), "Truth," in J. M. Baldwin (ed.), *Dictionary of Philosophy and Psychology*, vol. 2. New York: Macmillan, 716–18.
- Potter, M. (2009). *Wittgenstein's Notes on Logic*. Oxford: Oxford University Press.
- Quine, W. V. O. (1970). *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall.
- Ramsey, F. P. (1923). "Critical Notice" of Wittgenstein (1922), reprinted in Ramsey (1931), 270–86.
- Ramsey, F. P. (1927). "Facts and Propositions," reprinted in Ramsey (1931), 138–55.
- Ramsey, F. P. (1931), *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite. London: Routledge and Kegan Paul.
- Ricketts, T. (1996). "Pictures, Logic, and the Limits of Sense in Wittgenstein's *Tractatus*," in H. Sluga and D. G. Stern (eds.), *The Cambridge Companion to Wittgenstein*. Cambridge: Cambridge University Press, 59–99.
- Rumfit, I. (2011). "Ramsey on Truth and Meaning," in B. Morison and K. Ierodiakonou (eds.), *Episteme etc.: Essays in Honour of Johathan Barnes*. Oxford: Oxford University Press, 231–45.
- Russell, B. (1937) [1903]. *The Principles of Mathematics*, Cambridge: Cambridge University Press. 2nd edn (1937), London: George Allen & Unwin.
- Russell, B. (1904). "Meinong's Theory of Complexes and Assumptions," in *Collected Papers*, vol. 4, 432–74.
- Russell, B. (1905a). "On Denoting," in *Collected Papers*, vol. 4, 414–27.
- Russell, B. (1905b). "The Nature of Truth," in *Collected Papers*, vol. 4, 490–506.
- Russell, B. (1973) [1906]. "On the Substitutional Theory of Classes and Relations," in D. Lackey (ed.), *Essays in Analysis*. London: George Allen & Unwin, 165–89.
- Russell, B. (1906–7). "On the Nature of Truth," *Proceedings of the Aristotelian Society* 7: 28–49.
- Russell, B. (1910). "On the Nature of Truth and Falsehood," in *Collected Papers*, vol. 6, 115–24.
- Russell, B. (1911). "Knowledge by Acquaintance and Knowledge by Description," in *Collected Papers*, vol. 6, 147–61.
- Russell, B. (1912). *The Problems of Philosophy*. References to Home University Library edition.
- Russell, B. (1913). "Theory of Knowledge," in *Collected Papers*, vol. 7, 1–178.
- Russell, B. (1918). "The Philosophy of Logical Atomism," in *Collected Papers*, vol. 8, 157–244.
- Russell, B. (1921). *The Analysis of Mind*. London: George Allen & Unwin.
- Russell, B. (1922). "Introduction," in Wittgenstein (1922), 7–23.
- Russell, B. (1983–). *The Collected Papers of Bertrand Russell: The McMaster University Edition*, gen. ed. J. Passmore. London: Routledge.
- Strawson, P. (1950). "Truth," reprinted in P. Strawson, *Logico-Linguistic Papers* (1971). London: Methuen, 190–211.
- Whitehead, A. N. and B. Russell (1910). *Principia Mathematica*, vol. 1. Cambridge: Cambridge University Press (2nd edn, 1927).
- Wittgenstein, L. (1913). *Notes on Logic*, in Wittgenstein (1979a).

- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*, trans. C. K. Ogden. London: Routledge and Kegan Paul (2nd edn, 1933).
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Oxford: Blackwell.
- Wittgenstein, L. (1974). *Philosophical Grammar*, ed. R. Rhees, trans. A. Kenny. Oxford: Blackwell.
- Wittgenstein, L. (1979a). *Notebooks 1914–1916*, trans. G. E. M. Anscombe. Oxford: Blackwell.
- Wittgenstein, L. (1979b). *Wittgenstein's Lectures, Cambridge 1932–1935*, ed. A. Ambrose. Oxford: Blackwell.
- Wittgenstein, L. (2008). *Wittgenstein in Cambridge: Letters and Documents 1911–1951*, ed. B. F. McGuinness. Oxford: Blackwell.

CHAPTER 7

TRUTH IN FREGE

RICHARD KIMBERLY HECK AND ROBERT MAY

TARSKI is widely regarded as the father of formal semantics, and rightly so. But if Tarski is the father, then Frege is the grandfather. Frege's *Grundgesetze der Arithmetik* contains a semantic theory for his formal language, his "begriffsschrift" (or "conceptual notation"), that is no less rigorous than the one for the calculus of classes that Tarski develops in the early parts of "The Concept of Truth in Formalized Languages" (Tarski 1958). Like Frege's semantics for *begriffsschrift*, Tarski's is stated in an informal meta-theory and is in no sense "formal." Moreover, Frege argues in section 31 of *Grundgesetze* that his semantics is adequate to assign a unique denotation to every expression of *begriffsschrift*, and this argument has essentially the same purpose, and much the same structure, as Tarski's proof that his semantics is materially adequate.¹

Of course, there are differences between Frege and Tarski. Most of these derive from the fact that they have very different reasons for being interested in semantic theory. Frege wouldn't have seen much point in formalizing the semantics for his formal language,² since its purpose was not to report the pre-established meanings of the expressions of that language but rather to establish those meanings. Formalizing the semantic theory for *begriffsschrift* in *begriffsschrift* would therefore have been unhelpful. Tarski, on the other hand, is quite plainly interested in formalized semantic theories, and, if he does not present the semantics for the calculus of classes as a formal theory, then that is because it would have been obvious enough to his readers how it could be formalized. Indeed, Tarski would have expected his readers already to be familiar with the general structure of semantic theories, much as, in his paper on the completeness theorem, Gödel (1930) expected his readers to be familiar with the structure of models

¹ Unfortunately, space constraints require that much of what we say here be dogmatically asserted. For defense of this particular dogmatic assertion, see Heck's book *Reading Frege's Grundgesetze* (Heck 2012: Part I) and the papers on which that material was based (Heck 1998; 1999; 2010).

² Some commentators would argue that Frege would have found the very project of formal semantics unintelligible. This is often taken to be at the crux of his dispute with Hilbert. We disagree (Antonelli and May 2000), and our claim here is much weaker.

of first-order theories. Both expected their readers to be familiar with the technical background that made the discussion intelligible.

The later parts of “The Concept of Truth” are, of course, focused on formalized semantic theories, primarily on ones formalized in the theory of types. But Tarski’s central interest is not so much in the structure of semantic theories as it is in the question of whether such theories can be immunized from the semantic paradoxes. What made that issue pressing was the fact that, by the time Tarski did his work, the notion of truth had assumed a central place in logical theory; it is in terms of truth for example that logical validity is defined. Yet the notion of truth was subject to paradoxes that could not be ignored forever, especially after Gödel (1931) showed us, in his paper on the *incompleteness* theorems, that the self-reference needed for the paradoxes was available already in elementary arithmetic. For Tarski, then, formalizing semantic theories—and clarifying the semantics of quantification, via the notion of satisfaction—was just one step toward his real goal: showing how semantic theories can be consistently developed.

Frege, by contrast, never mentions the semantic paradoxes and was not interested in “theories of truth” in the way Tarski, Kripke (1975), and so many others since have been (this volume, *passim*). That is very much not to say that Frege was not interested in truth. Frege was profoundly interested in logic, and, on his view, truth is the real subject-matter of logic. As he puts it in a famous passage,

Just as “beautiful” points the way for aesthetics and “good” for ethics, so does the word “true” for logic. All sciences have truth as their goal; but logic is also concerned with it in a quite different way: logic has much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth. (Tht, op. 58)³

Given this conception of logic, it should be no surprise that, among the many tasks Frege sets himself in Part I of *Grundgesetze*, in which he explains his formal system, is that of proving that each of his “Basic Laws” is true and that each of his rules of inference preserves truth (Heck 2010; 2012: ch. 2). The semantic theory Frege gives for *begriffsschrift* plays a critical role in these arguments. It is because of the particular semantic clause that governs the conditional, for example, that Frege’s Basic Law I⁴

$$\vdash p \rightarrow (q \rightarrow p)$$

is true and that his first method of inference, *modus ponens*, is valid. Frege’s arguments for the truth of Law I (Gg, v. I, section 18) and for the validity of *modus ponens* (Gg, v. I, section 14) appeal directly, and only, to that clause.

³ References to Frege’s published papers are given by the page number in the original publication. Most reprintings include this information.

⁴ We shall silently translate Frege’s notation into modern notation, only preserving his prefixed vertical “judgement” and horizontal “content” strokes.

For Frege, then, truth is the most fundamental notion for both logic and semantics. This doctrine surfaces most directly in Frege's distinctive, and oft misunderstood, claim that the reference of a sentence is its truth-value. As we shall see, this claim has two complementary aspects: it underlies Frege's truth-functional understanding of the sentential operators, and it surfaces in his treatment of concepts as functions from objects to truth-values.⁵ We shall also discuss Frege's seemingly puzzling claim that the truth-values are objects, in an effort to make it somewhat less puzzling. We shall then turn our attention to the brief remarks Frege does make about the "truth-predicate." Finally, we shall consider, very briefly, the question of whether Frege was a deflationist.

7.1 SENTENCES REFER TO TRUTH-VALUES

As just said, Frege's most important thesis about truth is that the reference of a sentence is its truth-value. We shall call this the "Truth-Value Thesis."

Frege himself does not clearly distinguish the Truth-Value Thesis from the thesis that truth-values are *objects*. As Dummett has emphasized,⁶ since, for Frege, ontological categories supervene on syntactic ones, this latter thesis is all but equivalent to the claim that sentences are a sort of proper name. It should be obvious that this syntactic thesis is independent of the Truth-Value Thesis. We will discuss the claim that truth-values are objects in section 7.4.

As Dummett also makes clear, the claim that sentences *refer*, while it sounds odd to the modern ear, should not be doubted on that ground.⁷ Frege's notion of reference, though grounded intuitively in the relation between a name and its bearer, is a technical one, to be explicated in terms of the theoretical work it is supposed to do. That work is semantic. So the Truth-Value Thesis amounts to the claim that the fundamental semantic fact about a sentence is that it is true or false: that it has whatever truth-value it has. Dummett therefore suggests that we state Frege's view as: "The 'semantic value' of a sentence is its truth-value." This usage has become almost standard in truth-theoretic semantics (see e.g. Larson and Segal 1995). But we shall defer to Frege's usage and so will continue to speak of sentences as "referring" to truth-values.

It is a common complaint that Frege gives little direct argument for the Truth-Value Thesis. The arguments he does give are in "On Sense and Reference," and they are clearly inadequate. The first argument is that it only matters to us whether a name refers in so far as we are concerned with the truth-value of some sentence in which it occurs. That

⁵ Unfortunately, we shall not have space to discuss Frege's truth-conditional conception of meaning, which is another aspect of the thesis.

⁶ The various claims of Dummett's we shall mention come primarily from *Frege: Philosophy of Language* (Dummett 1981: esp. ch. 12).

⁷ Frege's German term is *bedeuten*, whose most natural translation is "mean." But that doesn't help, since the claim that sentences "mean" their truth-values sounds odd for a different reason.

may be true, but it hardly “drive[s us] into accepting the truth-value of a sentence as constituting what it refers to” (SM, op. 34).

A second set of considerations follows:

If our supposition that the reference of a sentence is its truth-value is correct, then the latter must remain unchanged when a part of the sentence is replaced by an expression with the same reference. And this is in fact the case . . . If we are dealing with sentences for which the meaning of their component parts is at all relevant, then what feature except the truth-value can be found that belongs to such sentences quite generally and remains unchanged by substitutions of the kind just mentioned? (SM, op. 35)

This looks like an attempt to derive the Truth-Value Thesis from the compositionality of reference, that is from the claim that the reference of a complex expression is determined by the references of its parts. But it too is unconvincing. What it shows is that it is *consistent* with compositionality to take sentences to refer to their truth-values. If Frege cannot think of another option, then that, one might think, is his problem.⁸

Why are Frege’s arguments for the Truth-Value Thesis so pathetic? The answer is simple: he doesn’t really have a direct argument for it. His argument is ultimately pragmatic, as he indicates in *Grundgesetze*: “How much simpler and sharper everything becomes with the introduction of truth-values, only detailed acquaintance with this book can show” (Gg, x). The Truth-Value Thesis solves a lot of problems, and it solves them, Frege thinks, better than anything else on offer. That is his real argument.

Given Frege’s commitment to the compositionality of reference, the Truth-Value Thesis has two complementary aspects, since sentences both *have parts* themselves and *are parts* of other sentences. Now, in Frege’s logic, the sentential operators are negation and the conditional. So compositionality and the Truth-Value Thesis imply that negation and the conditional are truth-functional. That is the first of the two complementary aspects. The other is the thesis that the reference (semantic value) of a predicate is a function from objects to truth-values. As Dummett again makes clear, this all but follows from the Truth-Value Thesis, given three of Frege’s other commitments: that the reference of a predicate is a function; that names refer to objects; and that the reference of a sentence “*Fa*” is the result of applying the function to which “*Fξ*” refers to the object to which “*a*” refers.⁹ For consider such a sentence. We know that “*Fξ*” must denote a function and that “*a*” must denote an object. So the arguments to the function “*Fξ*” are certainly objects, and if the reference of “*Fa*” is to be a truth-value, then the values of the function must be truth-values. So “*Fξ*” denotes a function from objects to truth-values. Such functions are what Frege calls “first-level concepts.”¹⁰

⁸ And, of course, there is another option: the Russellian proposition expressed by the sentence will remain unchanged.

⁹ This last doctrine reflects Frege’s view that semantic composition is “internal” to the semantics of predicates, in a sense we try to explain elsewhere (Heck and May 2006; 2013).

¹⁰ Frege holds similar views about predicates of other logical types: e.g. a two-place predicate like “*ξ loves η*” refers, on Frege’s view, to a two-place function from objects to truth-values—to a two-place,

In the presence of compositionality, the Truth-Value Thesis thus entails that sentential operators are truth-functional and that concepts are functions from objects to truth-values. And now the important observation is that *neither* of these claims is present in Frege's early work. Rather, they emerge as responses to problems Frege had, problems that result from the failure of an earlier view. So, to understand why Frege adopts the Truth-Value Thesis, we need to understand those problems and how the Truth-Value Thesis solves them.

7.2 CONCEPTS ARE FUNCTIONS FROM OBJECTS TO TRUTH-VALUES

Let us focus first on Frege's doctrine that concepts are functions from objects to truth-values. The idea that concepts are functions is already present in *Begriffsschrift*. The distinction between function and argument is there presented as the key to Frege's new analysis of generality, and Frege explains his distinction between function and argument in *Begriffsschrift* (Bg, section 9) almost exactly as he explains the distinction between concept and object in his mature work (FC, op. 6). Frege does not use the term "concept" in *Begriffsschrift*, but when he introduces the distinction between function and argument, he does so by applying it to the sentence "Hydrogen is lighter than carbon dioxide." Frege says that, if we regard "hydrogen" as replaceable by other expressions, then "'hydrogen' is the argument and 'being lighter than carbon dioxide' is the function . . ." (Bg, section 9). Obviously, "being lighter than carbon dioxide" is a predicate, and so the upshot is that we are to regard it, logically, as a function.

Frege thus regarded a simple sentence like "Bob runs" as being analyzable into a function, "runs," and an argument, "Bob." For us, the next question to ask would be: what are the inputs and outputs of this function? In *Begriffsschrift*, however, the question does not arise, since Frege's official view there is that functions are *expressions* (Bg, section 9). Frege abandons this view, however, by 1881 (Heck and May 2010; 2013), and then the question does arise as to what the arguments and values of these 'concept-functions' are. It is clear enough what the arguments must be: Frege held already in *Begriffsschrift* that the content of a proper name is the object it denotes (Bg, section 8), so the arguments are just ordinary objects. But what are the *values* of concept-functions supposed to be?

In his writings from the early 1880s, Frege does not explicitly answer this question, but it is easy enough to deduce the answer. Consider a simple sentence "*Fa*." Applying the concept-function that is the content of "*Fξ*" to the object that is the content of "*a*"

first-level concept. Higher-level concepts take concepts (more generally, functions) as arguments. Quite generally, then, predicates refer to functions from values of some appropriate type (or types) to truth-values.

should yield the content of the complex expression that they together constitute. But Frege's view in the early 1880s was that the content of "*Fa*" is a content "that can become a judgement" (Bg, section 2), for short, a judgeable content. So Frege's original view must have been the following: a proper name has as its content an object; a predicate has as its content a function from objects to judgeable contents; and the content of "*Fa*," which is a content that can become a judgment, is the result of applying the function that is the content of "*F*" to the object that is the content of "*a*."¹¹

This view is really quite elegant, not to mention strikingly reminiscent of Russell's later notion of a propositional function (Russell 1903: ch. 7). Unfortunately, as a now famous argument shows, it has consequences that would have been unacceptable to Frege. If the content of

- (1) The Evening Star is a planet

is the result of applying the function that is the content of "ξ is a planet" to the object that is the content of "the Evening Star," then

- (2) The Morning Star is a planet

must have the same content. But (1) and (2) cannot have the same content, not if identity of content is to guarantee identity of logical properties, as Frege clearly insists it must.¹²

[T]he contents of two judgements may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgements, always follow from the second, when it is combined with these same judgements, or this is not the case . . . I call that part of the content that is the *same* in both the *conceptual content* . . . Everything necessary for a correct inference is expressed in full, but what is not necessary is generally not indicated; *nothing is left to guesswork*. (Bg, section 3; emphasis in original)

And now, since (1) and (2) have different logical consequences, they must have different contents.

How can Frege avoid the conclusion that (1) and (2) must have the same content? The conclusion depends upon just four assumptions:

- (i) The content of a proper name is its bearer.
- (ii) Concepts are functions.

¹¹ Beaney (2007) comes to the same conclusion.

¹² Some readers—e.g. Brandom (1994: 94ff) and Kremer (2010)—have claimed to find a stronger thesis here: that, if *A* and *B* have the same consequences, then they must also have the same content. We don't read the passage that way, but the issue is not relevant at present.

- (iii) The content of a simple sentence "*Fa*" is the result of applying the concept-function that is the content of "*F*" to the content of "*a*."
- (iv) Logical properties are determined by content, so that sameness of content implies sameness of logical properties.

Only (iii) is dispensable: (ii) is the key to Frege's understanding of logical generality; (iv) is integral to Frege's conception of logic and its relation to content (May 2006); (i) is central to Frege's understanding of identity as objectual. This last view is not present in *Begriffsschrift* but is in place by 1881 and is fundamental to Frege's logicism (May 2001; Heck and May 2006; Heck 2015). So, as said, (iii) is what Frege must abandon: the content of "*Fa*" *cannot* be the result of applying the concept-function that is the content of "*F*" to the object that is the content of "*a*." To put it differently: the values of concept-functions *cannot* be judgeable contents. So what are they?

Frege's mature view, of course, is that the values of concept-functions are truth-values: concepts are functions from objects to truth-values. This thesis is thus partly a result of Frege's being forced to re-think the question of what the values of concept-functions are, once he realizes that they cannot be judgeable contents.

One source of Frege's new view is how he proposes to distinguish concepts from objects. In a letter to the philosopher Anton Marty, written in 1882, Frege remarks that he "regard[s] it as essential for a concept that the question whether something falls under it has a sense," a question that would be senseless in the case of an object (PMC, 101). The view that concepts are functions from objects to truth-values fits naturally with this suggestion, since what such a function does is sort objects into two baskets: those that fall under the concept and those that do not.

Such a consideration does make the view that the truth-values are the values of concept-functions attractive, but more would have been needed to drive Frege to it, because it has deeply counterintuitive consequences.¹³ In particular, it implies that concepts are extensional, so that there can be only one concept true of a given collection of objects. For example, if, as philosophers' lore has it, the animals that are supposed to have kidneys are the same as the animals that are supposed to have hearts, then the concepts *renate* and *cordate* will be the same concept, and that has no intuitive plausibility. Of course, the mature Frege does think that concepts are extensional. But that was not always his view, either.

The view that concepts are intensional is an almost immediate consequence of Frege's early view that concepts are functions from objects to judgeable contents. On that view, *is a renate* maps Bob to *Bob is a renate*, and *is a cordate* maps Bob to *Bob is a cordate*. Since the content *Bob is a cordate* has different logical properties from the content *Bob is a renate*, these are different contents, and so the functions *is a cordate* and *is a renate* are distinct: they have different values for the argument Bob and, indeed, for every

¹³ The argument to be given depends upon the assumption that functions themselves are extensional. We discuss this matter elsewhere (Heck and May 2010: section 2).

argument. Moreover, in 1881, Frege is prepared to use the thesis that concepts are intensional as a premise—without saying even a word in defense of it—in one of his most important arguments, an argument for the conclusion that concepts must be distinguished from objects (BLC, 18). And yet, by 1884, Frege has changed his mind and come to regard concepts as extensional (Gl, section 68 fn. 1). It is hard to imagine any explanation for this shift other than that Frege has, in the intervening years, changed his mind about what the values of concept-functions are: not judgeable contents, but truth-values.

So why that change? To answer this question, we need to look at the other aspect of the Truth-Value Thesis, the one connected with the fact that sentences can occur as parts of other sentences.

7.3 SENTENTIAL CONNECTIVES AS TRUTH-FUNCTIONS

Frege has sometimes been credited with the discovery of truth-tables (Kneale and Kneale 1962: 420, 531; Wittgenstein 1979b: 135ff), and something akin to truth-tables is indeed present in Frege's early work. Frege emphasizes that, when we are considering a binary sentential connective, we must distinguish four possible cases. But he does not mention the notion of truth in this connection: he does not say, in *Begriffsschrift*, that a conditional is false only when its antecedent is true and its consequent is false. Frege's explanation of the conditional reads, rather, as follows:

If *A* and *B* stand for contents that can become judgements . . . there are the following four possibilities:

- (1) *A* is affirmed and *B* is affirmed;
- (2) *A* is affirmed and *B* is denied;
- (3) *A* is denied and *B* is affirmed;
- (4) *A* is denied and *B* is denied.

Now $\vdash B \rightarrow A$ stands for the judgement that *the third of these possibilities does not take place, but one of the other three does*. (Bg, section 5, emphasis in original; see also BLC, 35)

Frege does not speak of truth in his explanation of negation, either (Bg, section 7). The suggestion that negation and the conditional express truth-functions is therefore wholly absent from *Begriffsschrift* and, indeed, from all of Frege's early writings.

Truth-functions appear in Frege's work for the first time in *Function and Concept*, published in 1891. After explaining his conception of functions as "incomplete," Frege motivates his view that concepts are functions whose values are always truth-values (FC, opp. 13ff). But if the truth-values are admitted as values of functions, we can allow them to occur as arguments, too. Interesting cases then include negation and the conditional (FC, opp. 20ff), and Frege clearly and explicitly explains these as truth-functions.

That said, so far as we know, nowhere in his later writings does Frege give the sort of “tabular” account that Wittgenstein and the Kneales mention, so there is no real basis for attributing the discovery of truth-tables to Frege.¹⁴

Perhaps surprisingly, even the idea that the sentential operators express functions seems to be absent from *Begriffsschrift*. Its emergence is due to Frege’s belated encounter with the work of George Boole, whom Frege seems to have read only after the appearance of scathing reviews of *Begriffsschrift* published by leading members of the Boolean school. Both Ernst Schröder (1972) and John Venn (1972) accuse Frege of essentially replicating Boole’s work, but in a less satisfactory form and only partially.

Frege criticizes Boolean logic on several grounds. Perhaps his most penetrating criticism concerns Boole’s distinction between the “calculus of judgements” and the “calculus of concepts.” The former is essentially what we know as propositional logic; the latter treats of relations between concepts, those expressed by “All *F* are *G*,” “some *F* are *G*,” and so forth. Boole regards the calculus of concepts as basic and reduces the calculus of judgments to it. Frege argues against Boole on both points: he claims that Boole’s attempt to reduce the calculus of judgments to the calculus of concepts is a failure and that Boole is wrong to treat the latter as basic.¹⁵

Boole’s reduction proceeds as follows. Both calculi contain expressions of the forms “ $A \times B$,” “ $A + B$,” “ \bar{A} ,” and so forth.¹⁶ In the calculus of concepts, the letters are taken to denote classes (or extensions of concepts), and the operations are interpreted set-theoretically: multiplication is intersection; addition is union; the bar represents complementation relative to the chosen “universe of discourse.” How are the letters interpreted in the calculus of judgments? One might expect that they would denote truth-values, but that would be wrong.¹⁷ They again denote *classes*, so the operations are still set-theoretic. In *The Laws of Thought*, for example, Boole takes the letters in the calculus of judgments to denote classes of times: the times a proposition is true. So a conditional (judgment) becomes a universal affirmative proposition (relation between classes): all times at which the antecedent is true are times the consequent is true (Boole 1854: ch. 11, section 5).

¹⁴ Moreover, Frege never considers truth-tables for arbitrary formulae (e.g. $p \vee r \rightarrow q \vee r$), but only for the simplest cases, and there is no indication that he realized, as both Wittgenstein and Post (1921) did, that truth-tables can be used to determine the validity of an arbitrary propositional formula. As is now widely recognized, then, it is Wittgenstein and Post who deserve the real credit for the discovery of truth-tables.

¹⁵ Frege also points out that Boole has no way to mix the two, as in: $\forall x(Fx \rightarrow Gx) \vee \exists x(Gx \wedge Hx)$. And, of course, that Boole cannot handle nested quantifiers, as in: $\exists x \forall y(Rxy)$.

¹⁶ The notation varies from logician to logician, as do the details of its interpretation, but these differences do not matter for our purposes.

¹⁷ MacColl (1877: 9–10) comes closest to this conception, but his official view is that the sentence-letters denote “statements” Schröder mentions MacColl in his review, but it is unclear if Frege ever read him. Frege mentions MacColl twice (AimCN, 93; BLF, 15), but what he says is all but lifted from Schröder.

Boole's central idea, of course, is to treat the sentential operators as expressing set-theoretic operations on the power set of some universe. What the universe comprises in the case of judgments varies, both in Boole's own work and in that of his followers, but that has proved not to be the crucial point. The crucial point is that, once we treat the sentential operators set-theoretically, the algebra so determined is (what we now call) a Boolean algebra, and it validates the laws of classical logic. Sadly, Frege could no more see the importance of this idea than Schröder could see the importance of Frege's new analysis of generality. But Frege is surely right that the attempted reduction of sentential logic to quantification theory is a failure, and not only for the case of "eternal truths such as those of mathematics" (BLC, 15). What is fundamental is sentential logic, and Frege goes so far as to describe himself, somewhat misleadingly, as reducing universal affirmative propositions to conditionals (BLC, 17–18).

There is much more to be said about this, but just two points are important here. The first is that Boole does *not* treat the sentential connectives as expressing *truth*-functions. Boole does regard the case in which the universe contains just one element as important, but this is mostly because it makes the calculations in which he is interested especially simple. And the two elements of the algebra so determined, though denoted "1" and "0," are not the True and the False. As always in Boole, "1" denotes the universe, and "0" denotes the empty set, and the operations are set-theoretic.

The second point is that Boole undoubtedly *did* treat the sentential connectives as expressing *functions*. His use of the arithmetical expressions "+" and "×" serves to emphasize this point. And, as critical as Frege is of Boole's over-reliance on the similarities between logic and arithmetic (AimCN, 93–4; BLC, 13–15), it is hard to imagine that Frege would not have been struck by this element of Boole's work. It might even have seemed like confirmation of his own emphasis on the importance of the notion of *function* to logic. Frege does not highlight this aspect of Boole's work in his critical pieces—he is too busy defending himself—but nor does he criticize it. So, although it may have taken him a little while to assimilate it, it seems very plausible that Frege got the idea that the sentential connectives express functions from his reading of Boole.

So we ask: what are the arguments and values of these functions? At one time, of course, the obvious thing for Frege to say would have been that the conditional expresses a two-place function from contents to contents. But, once Frege discovers the substitution argument, this option is off the table, since sentences cannot have judgeable contents as their semantic values.

So what are the arguments and values of the conditional, if they are not judgeable contents? The answer is almost there in *Begriffsschrift*. As we saw earlier, Frege distinguishes four possibilities:

- (1) A is affirmed and B is affirmed
- (2) A is affirmed and B is denied
- (3) A is denied and B is affirmed
- (4) A is denied and B is denied

and says that “ $\vdash B \rightarrow A$ stands for the judgement that the third of these possibilities does not take place” (Bg, section 5). This, of course, is wrong.

The language of affirmation and denial is not only quaint but misplaced, as Frege himself would eventually come to realize. This is essentially what Peter Geach (1965: 449) famously called “the Frege point.” It is closely connected with what Frege himself called “the dissociation of assertoric force from the predicate” and regarded as one of his most important discoveries (WMR): if one asserts a conditional, then it is only the conditional as a whole that is affirmed; one need neither deny its antecedent nor affirm its consequent.¹⁸ These contents are *judgeable*, not *judged*. *Contra* Geach, however, the Frege point is not really there in *Begriffsschrift*. There are intimations of it, to be sure (Bg, section 2). But if Frege had fully appreciated it, he could not have used the language of affirmation and denial the way he does in explaining the conditional.

The language of affirmation and denial is present in Frege’s writings through 1882; it does not appear later. We speculate, therefore, that at some point between 1882 and 1884, Frege must have become dissatisfied with his use of such language, presumably for the reason we have just recalled. When he does, he needs a new way to explain the conditional; he needs to reconstruct his table of possibilities in terms of something other than affirmation and denial. In the writings from 1881 and 1882, Frege sometimes presents the table this way:

- (1) A and B
- (2) A and not- B
- (3) not- A and B
- (4) not- A and not- B

remarking that his “ $B \rightarrow A$ denies the third” (BLC, 35). The language of affirmation and denial is thus almost gone. But this form of explanation will still be inadequate once Frege has decided that the conditional ought to be understood as expressing a function: we need to know what the arguments and values of that function are.

With this question in the front of his mind, and Boole’s work in the back of his mind, Frege was eventually struck by an answer: what matters is not whether A and B are affirmed or denied but just whether they are true or false. And now the other piece falls into place: if the arguments and values of the connectives are truth-values, then the values of concept-functions must be truth-values, as well, since the values of concept-functions are also arguments of the connectives.

No doubt, there are other options. But that is why we said, earlier, that Frege’s argument for the Truth-Value Thesis is ultimately pragmatic. Frege had a whole set of problems that were generated by the failure of his original view that sentences refer to

¹⁸ Similarly, Frege’s explanation of the conditional is insufficiently general, since it tells us only about *asserted* conditionals and does not seem to apply to embedded conditionals.

judgeable contents. The Truth-Value Thesis not only solves those problems, it does so in an elegant way, not just tying the loose ends but tying them neatly. So now we can understand what Frege meant when he spoke of “[h]ow much simpler and sharper everything becomes with the introduction of truth-values” (Gg, x). Though he might have added, “It would help to look at my earlier work to get a sense of what a mess things were before.”

Of course, Frege’s new view gives rise to problems of its own. We have already mentioned one counterintuitive consequence: the extensionality of concepts. But there is an even worse problem:¹⁹

[W]hat “ $2^2 = 4$ ” means is the True just as, say, “ 2^2 ” means 4. And “ $2^2 = 1$ ” means the False. Accordingly, “ $2^2 = 4$,” “ $2 > 1$,” and “ $2^4 = 4^2$ ” all mean the same thing, viz., the True . . . The objection here suggests itself that “ $2^2 = 4$ ” and “ $2 > 1$ ” nevertheless tell us quite different things, express quite different thoughts. (FC, 13)

If now the truth-value of a sentence is its meaning, then on the one hand all true sentences have the same meaning and so, on the other hand, do all false sentences. From this we see that in the meaning of the sentence all that is specific is obliterated. (SM, 35)

7.4 TRUTH-VALUES AS OBJECTS

We have argued that Frege’s mature view that sentences refer to truth-values emerged between 1881 and 1884, when he was engaged with the work of the Booleans. It is important to understand that we are *not* claiming that Frege held already in 1884 that truth-values are *objects*.²⁰ As already noted, the two views are distinct. The ontological thesis that truth-values are objects is bound up with the syntactic thesis that sentences are proper names, and there is reason to believe that Frege had not arrived at this latter view by 1884. In particular, there are significant differences between his statements of the “context principle” in *Die Grundlagen* (Gl, x and section 62) and in *Grundgesetze* (Gg, v. I, section 29). As Dummett (1981: ch. 12) has argued, these differences reflect a change from a view on which sentences have a privileged position to one on which they are but a kind of proper name (their syntactic complexity notwithstanding). It is anyway reasonable to suppose that it took some time for Frege fully to adjust his conception of logic to his discovery of the truth-values. The puzzle mentioned at the end of the last section, for example, was one to which Frege did not have a satisfactory response until “On Sense and Reference,” published in 1891 (Heck and May 2010).

¹⁹ We are here translating Frege’s German term “*bedeuten*” (and its cognates) using the ordinary English equivalent, “mean,” as one simply cannot appreciate the force of the objections Frege is considering otherwise.

²⁰ So we could agree with Goren Sundholm (2001: 62) that “Frege did not have the doctrine of (objectual) truth-values” until about 1890. Whether that is the right thing to say, however, is not so clear (Heck 2015: section 2).

One might think, however, that Frege already treats sentences as proper names in *Begriffsschrift*. Just as in *Grundgesetze*, sentences occur in *Begriffsschrift* in positions where proper names occur. Frege is as happy to write

$$\vdash p \equiv \neg \neg p$$

as he is to write

$$\vdash A \equiv B,$$

where A and B are names of geometrical points, as in the famous example in *Begriffsschrift*, section 8. Similarly, the variables “ c ” and “ d ” that occur in Frege’s proposition (57)

$$\vdash c \equiv d \rightarrow (Fd \rightarrow Fc)$$

may be replaced either by singular terms or by sentences. For example, the proof of proposition (68)

$$\vdash \forall a(f(a)) \equiv b \rightarrow (b \rightarrow f(c))$$

requires the substitution of “ $\forall a(Fa)$ ” for “ c ” in (57).

There is nonetheless a different sort of distinction between names and sentences in *Begriffsschrift*. Frege insists that “[w]hatever follows the content stroke must have a content that can become a judgement” (Bg, section 2). So one cannot write “— A ” unless A is a formula. Thus, in (68), b must be a formula, since it appears as antecedent of a conditional and so is preceded by the content stroke (as is apparent when the formula is given in Frege’s notation). And, though Frege does not explicitly exclude such constructions, there is no evidence in *Begriffsschrift* that he would have regarded something with a sentence on one side of “ \equiv ” and a name on the other, such as

$$\forall a(Fa) \equiv 1,$$

as well-formed. Matters are very different in *Grundgesetze*, where

$$\forall a(a = a) = \epsilon'(-\epsilon),$$

which says that the True is the extension of the concept *is identical with the True*, is not just well-formed but true.

Why then does Frege treat the truth-values as objects?²¹ It is often suggested that he has no choice but to do so, since, on his view, there are only two kinds of “things,” the complete and the incomplete. Objects are the paradigmatic complete entities; concepts

²¹ It might be thought that there was a prior question, namely, why Frege treats truth-values as any kind of “thing.” But this has already been answered: if concepts are to be functions from objects to truth-values, and if sentential connectives are to be truth-functional, then truth-values have to be able to occur as arguments and values of functions. So they have to be some kind of “thing,” and that view has to be in place by 1884, since it is what drives the view that concepts are extensional.

and functions the incomplete ones. Since the truth-values are not incomplete, they must be complete, and so must be objects. But this line of reasoning is at best inconclusive. There is no obvious reason Frege could not have held that there was more than one kind of complete entity.²² Indeed, since, for Frege, ontological categories largely supervene on syntactic ones, the apparent grammatical differences between sentences and names speak strongly in favor of such a distinction.

If there were more than one type of complete entity, then the logic would become, in present-day terminology, “many-sorted.” There would then be four types of one-place, first-level functions: from objects to objects; from objects to truth-values; from truth-values to objects; and from truth-values to truth-values. Similarly, there would be two different notions of identity. But we already have such distinctions in the hierarchy of types: Frege insists, for example, that we cannot really say that two functions are “identical,” since identity is a relation between objects (CSM, 121).

To understand why Frege treats truth-values as objects, and what his so treating them amounts to, we need to look closely at a particular feature of the formal development in *Grundgesetze*.

Value-ranges are to functions as extensions are to concepts; think of a function's value-range as its graph (in the set-theoretic sense). Frege's Basic Law V governs the notation for value-ranges:

$$\vdash (\dot{\epsilon}(f\epsilon) = \dot{\epsilon}(g\epsilon)) = \forall x(fx = gx)$$

This says that the value-range of the function $f\xi$ is the same as that of $g\xi$ just in case they always have the same value for the same argument. One might have thought that this would only allow Frege to speak of the value-ranges of one-place functions and that he would need similar axioms for two-place functions and the like. But one of the most elegant features of the formal development in *Grundgesetze* is how Frege handles the value-ranges of two-place functions, namely, as what he calls “double value-ranges.” Consider for example the function $\xi + \eta$. Fix its second argument and consider (e.g.) the function $\xi + 2$. The value-range of this function, $\dot{\epsilon}(\epsilon + 2)$, is the graph of the function whose value, for a given argument x , is $x + 2$. Suppose we now allow the second argument in $\dot{\epsilon}(\epsilon + 2)$ to vary; the resulting function, $\dot{\epsilon}(\epsilon + \eta)$, maps a given argument y to the value-range $\dot{\epsilon}(\epsilon + y)$. So what is the value-range of the function $\dot{\epsilon}(\epsilon + \eta)$? It is the *double* value-range $\dot{\alpha}\dot{\epsilon}(\epsilon + \alpha)$, the graph of the function whose value, for argument y , is the value-range $\dot{\epsilon}(\epsilon + y)$.²³ The analogue of Law V for two-place functions then emerges as a theorem:²⁴

$$\vdash [\dot{\alpha}\dot{\epsilon}(f\epsilon\alpha) = \dot{\alpha}\dot{\epsilon}(g\epsilon\alpha)] = \forall x\forall y(fxy = gxy)$$

²² Similar issues arise in connection with Frege's claim that numbers are objects (Heck 2011).

²³ Frege defines ordered pairs in terms of value-ranges, so value-ranges are not sets of ordered pairs. But pretend they are. Then the value-range of $\xi + 2$ is the set of ordered pairs $\{ \langle \epsilon, \epsilon + 2 \rangle \}$. The value-range of $\epsilon + y$ is the set of ordered pairs $\{ \langle \epsilon, \epsilon + y \rangle \}$. The value-range of the function $\dot{\epsilon}(\epsilon + \eta)$ is the set of ordered pairs $\{ \langle \alpha, \dot{\epsilon}(\epsilon + \alpha) \rangle \}$, or $\{ \langle \alpha, \{ \langle \epsilon, \epsilon + \alpha \rangle \} \rangle \}$, which we might then identify with the set of ordered triples: $\{ \langle \alpha, \epsilon, \epsilon + \alpha \rangle \}$.

²⁴ Frege does not prove this result, since he does not need it in this form. Theorems 2 and 3 do the necessary work.

And, of course, the same construction allows Frege to use the double value-range $\hat{\alpha}\hat{\epsilon}(R\epsilon\alpha)$ of a relation as its extension: extensions, quite generally, are just the value-ranges of concepts, i.e. of functions from objects to truth-values.

Nice, isn't it? But this trick depends upon Frege's identification of truth-values as objects. Suppose we do not treat truth-values as objects but as complete entities of a different sort. Then we need some other notation for extensions, say, " $\hat{x}(Fx)$," and a new Law governing that notation, say:²⁵

$$\hat{x}(Fx) = \hat{x}(Gx) \equiv \forall x(Fx \equiv Gx)$$

Then the "double extension" term " $\hat{y}\hat{x}(x < y)$," which one might have supposed would denote the extension of the relation $\xi < \eta$, is not even well-formed. Extension terms are formed by prefixing " \hat{y} ," say, to a one-place *predicate*, but " $\hat{x}(x < \eta)$ " is a functional expression, not a predicate: it denotes a function from objects to extensions.

The ugly workaround is to take the extension of a two-place predicate to be the value-range of a certain function from objects to extensions, i.e. $\hat{\alpha}\hat{x}(x < \alpha)$. But there is a more elegant solution. Select two arbitrary objects, which we shall denote " \perp " and " \top ," and suppose we can show that, for any concept F , there is a function whose value is \top for objects that fall under F and \perp for objects that do not.²⁶ We call this function the concept's *characteristic function*.²⁷ Then the extension of a concept may be taken to be the value-range of its characteristic function. It is easy to see that, so defined, extensions satisfy the law governing them displayed above.

In fact, one can almost get by with nothing but characteristic functions. Instead of a relation of identity, for example, we could make use only of its characteristic function. An expression like " $2 + 2 = 4$ " would then be a name of \top , not a sentence. And we can do the same with the logical constants: so " $(2 + 2 = 4) \wedge (1 + 1 = 3)$ " is a name of \perp . To be able to form sentences, and so make assertions, we would need to have at least one real predicate in the language. The most natural choice for such a predicate would be one which meant: ξ is identical with the True. This, of course, is essentially Frege's horizontal:²⁸ in *Grundgesetze*, — ξ is a function whose value is the True for the True as argument, and the False otherwise.

²⁵ Here " \equiv " is the distinct notion of identity for truth-values.

²⁶ One way to do this is to use a description operator. Frege has such an operator in the formal system of *Grundgesetze*, but it is applied to value-range terms, not to predicates.

²⁷ Nowadays, the characteristic function associated with a set S is defined as:

$$\phi_S(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{if } x \notin S \end{cases}$$

We have been unable to identify a clear antecedent of this notion prior to Frege's work, but do not know that there is not one.

²⁸ As we shall see in the next section, Frege did at one time hold that his *begriffsschrift* had only one real predicate.

But if we have come this far, we might well go a step farther and *identify* concepts with their characteristic functions. Once the identification is made, the truth-values simply become the objects \top and \perp , in terms of which the characteristic functions are defined, and “ $2 + 2 = 4$ ” is both a name of \top *and* a sentence. Nothing could be more natural, mathematically speaking. Treating sentences as being of the same logical type as proper names thus has substantial technical advantages in the context of Frege’s system. And once one sees that it amounts simply to identifying concepts with their characteristic functions, it should not seem all that perplexing.

If the truth-values are objects, however, then the question arises as to which objects they are; in particular, whether they might be objects we already know by other names. Frege addresses this question in section 10 of *Grundgesetze*.²⁹ He first argues that the stipulations he has made about the references of the primitive expressions of his language do not determine which objects the truth-values are. More precisely, he argues that those stipulations do not determine whether the truth-values are value-ranges and, if so, of which functions. He then stipulates that the truth-values are to be identified with their own unit classes: the True is identified with the value-range of the concept that maps only the True to the True; the False, with that of the concept that maps only the False to the True. Burge (2005) has argued that this particular identification was required by Frege’s other views, but we do not find his arguments convincing, in part because Frege seems to feel no need at all to justify the identification. Rather, Frege writes as if any stipulation at all will do. And if the interpretation given above is correct—if the thesis that truth-values are objects is really just an identification of concepts with their characteristic functions—then we can see why Frege feels free to stipulate which objects the truth-values are, so long as they are distinct. Why Frege thinks he needs to make any such stipulation is itself an interesting question, one far too large to discuss here. But making it the way he does must ultimately lead to disaster, since it obviously depends upon the inconsistent Basic Law V.

7.5 THE REGRESS ARGUMENT: WHY TRUTH IS NOT A PREDICATE

Much of the existing discussion of Frege’s views on truth focuses on an argument he gives for the conclusion that truth is indefinable. This argument, which has come to be known as the “regress argument,” appears in at least two places: the late essay, “Thoughts” (Tht, op. 60), written in 1917 or so, and an unfinished essay, “Logic,” which the editors of the *Nachlass* date to 1897. Here is the argument as it occurs in “Logic”:

Now it would be futile to employ a definition in order to make it clearer what is to be understood by “true”. If, for example, we wished to say that “an idea is true if it agrees

²⁹ Section 10 is not really about the truth-values but about an issue concerning value-ranges. There is an extensive literature on it.

with reality”; nothing would have been achieved, since in order to apply this definition we should have to decide whether some idea or other did agree with reality. Thus we should have to presuppose the very thing that is being defined. The same would hold good of any definition of the form “A is true” if and only if it has such-and-such properties or stands in such-and-such a relation to such-and-such a thing. In each case in hand it would always come back to the question whether it is true that A has such-and-such properties, stands in such-and-such a relation to such-and-such a thing. Truth is obviously something so primitive and simple that it is not possible to reduce it to anything still simpler. (Log97, 128–9)

This argument is extremely puzzling. Some commentators have found in it an argument that there is no real property of truth at all (Ricketts 1986; Kemp 1995). But one would need a very strong argument indeed for this sort of claim, since there are so many places in Frege’s writings where he seems to make serious use of semantic notions like reference and truth (Heck 2010; 2012: ch. 2). And, for the sorts of reasons given by Stanley (1996), Tappenden (1997), and Sullivan (2005), we do not think a strong argument has been given.

Perhaps what is most puzzling about the regress argument is that its first part contains what looks like a non-sequitur. Where does the question “whether some idea or other [does] agree with reality” presuppose the notion of truth? The later parts of the argument read differently. There Frege insists that the real question must be whether it is *true* that the idea agrees with reality. But that just seems gratuitous. Surely it is possible to ask whether Bob is home without asking whether it is true that Bob is home, let alone whether it is true that it is true that Bob is home, and so on and so forth, *ad infinitum*.

The key to Frege’s thinking is revealed by remarks that follow the regress argument proper:

What, in the first place, distinguishes [“true”] from all other predicates is that predicating it is always included in predicating anything whatever. If I assert that the sum of 2 and 3 is 5, then I thereby assert that it is true that 2 and 3 make 5. So I assert that it is true that my idea of Cologne cathedral agrees with reality, if I assert that it agrees with reality. Therefore, it is really by using the form of an assertoric sentence that we assert truth, and to do this we do not need the word “true”. (Log97, 129)

What lies at the core of the regress argument, then, is the idea that every assertion is an assertion of truth; every judgment, a judgment of truth.

Frege makes this sort of claim in many places, such as in “On Sense and Reference,” where he writes, “A judgement, for me, is not the mere grasping of a thought, but the admission of its truth” (SM, op. 34, note). This sort of perspective is critical to Frege’s larger conception of logic. Frege frequently emphasizes that his goal was “not . . . to present an abstract logic in formulas, but to express a content through written symbols in a more precise and perspicuous way than is possible with words” (AimCN, 90–1; see also Geo1; Geo2). More precisely, Frege’s logic was to be one we can actually use in reasoning,

in proving theorems, that is, where theorems are true contents. So logic, in that sense, issues in judgments, in “admissions of truth.”

The importance of this idea for the regress argument is clearest in an earlier presentation of essentially the same line of thought, again in “On Sense and Reference”:

One might be tempted to regard the relation of the thought to the True not as that of sense to reference, but rather as that of subject to predicate. One can indeed say: “The thought that 5 is a prime number is true”. But closer examination shows that nothing more has been said than in the simple sentence “5 is a prime number”. The truth claim arises in each case from the form of the assertoric sentence, and when the latter lacks its usual force, e.g., in the mouth of an actor upon the stage, even the sentence “The thought that 5 is a prime number is true” contains only a thought, and indeed the same thought as the simple “5 is a prime number”. It follows that the relation of the thought to the True may not be compared with that of subject to predicate. (SM, op. 34)

Here, Frege is arguing that predication of truth is inadequate for assertion. Rather, Frege says again, the act of assertion is effected by using a sentence of a certain form.³⁰ But perhaps what is most noteworthy is the context in which this argument occurs. Frege is arguing against the view that truth is a property of thoughts because he sees it as a competitor to his view that “the relation of the thought to the True [is] that of sense to reference.”

So we can reconstruct the regress argument as follows. Suppose we do think of truth as a property of thoughts (or propositions, or sentences, or what have you). Then the idea that judgment is admission of a thought’s truth becomes the idea that judging is predicating truth of a thought. What the regress argument shows is that this cannot be right. Predication, in this sense, is a sort of judgment: to predicate truth of the thought that p is just to judge that the thought that p is true, i.e. to judge that it is true that p . But then, to judge that it is true that p is to predicate truth of the thought that it is true that p , i.e. to judge that it is true that it is true that p . The regress is vicious, since the sense in which judgment is predication of truth was meant to be constitutive.³¹ So it is not just, as Frege puts it in “Logic,” that, to “assert truth, . . . we do not need the word ‘true.’” Rather, the right conclusion to draw is the one from “On Sense and Reference”: the word “true” *cannot* be used to assert truth.

What the regress argument shows, then, is that, while judgment does in some sense involve the acceptance of thoughts as true, to judge is not to predicate truth of a thought. We can therefore see why, as mentioned earlier, some commentators have been tempted to read Frege’s argument as showing that there is no such property of thoughts as truth.

³⁰ Davidson (1984) famously argues that the assertoric act is not conventional in the sense Frege seems to think it is. We think it unclear whether Frege is committed to such a view. His main point is the negative one: that predication of truth is inadequate for assertion.

³¹ It would not be vicious if the claim were, say, merely one about the commitments one incurs by making a judgment, as Dummett (1981: ch. 13) makes clear.

That is almost its conclusion. The real conclusion, however, is that truth is not *fundamentally* a property of thoughts: the role truth plays in judgment—what we gesture at when we say that judgment involves recognition of a thought as true—is more basic and direct than that. To secure that role, Frege claims, we must instead conceive of the relation between a thought and its truth-value as that of sense to reference. But that does not prevent there from being a property had by all and only those thoughts that refer to the True. Indeed, we have just said which property it is.

Further confirmation of this interpretation of the regress argument emerges if we ask who its target is. Who ever thought that judgment is predicating truth of a thought? Frege himself! Consider the following passage from *Begriffsschrift*:

We can imagine a language in which the proposition “Archimedes perished at the capture of Syracuse” would be expressed thus: “The violent death of Archimedes at the capture of Syracuse is a fact.” To be sure, one can distinguish between subject and predicate here, too, if one wishes to do so, but the subject contains the whole content, and the predicate serves only to turn the content into a judgement. *Such a language would have only a single predicate for all judgements, namely, “is a fact” . . . Our begriffsschrift is a language of this sort, and in it the sign ⊢ is the common predicate for all judgements.* (Bg, section 3; emphasis in original)

These remarks come at the conclusion of Frege’s explanation of why the “distinction between subject and predicate does not occur in [his] way of representing a judgement” (Bg, section 3). It is tempting, therefore, to regard them as but a grudging concession to tradition. But Frege emphasizes the final sentence of the quoted passage, and this sort of emphasis is used consistently in Part I of *Begriffsschrift* when he is articulating the central features of his new conception of logic. He is saying here, quite explicitly, that his *begriffsschrift* is a language in which there is only one predicate, the assertion-sign. What is most striking is his remark that “the predicate [‘is a fact’] serves . . . to turn the content into a judgement.” This looks like an explicit claim that assertion is achieved through the predication of facthood.³²

Frege thus held, throughout his career, that there is an intimate relationship between judgment and truth. But his original conception of this relation—that judgment is predication of truth—is unsatisfactory, because it falls to the regress argument. His mature view is that the relation between a sentence and its truth-value is to be modeled on the relation between a name and its bearer. To judge that 5 is prime is thus not to predicate truth of the thought that 5 is prime. It is to entertain the thought that 5 is prime in an attempt to refer to the True,³³ thus taking “the step from the level of thoughts to the level of reference”

³² As Proops (1997) points out, Wittgenstein ascribes this view to Frege: “The verb of a proposition is not ‘is true’ or ‘is false,’ as Frege thought; rather, that which ‘is true’ must already contain the verb” (Wittgenstein 1961: 4.063; see also Wittgenstein 1979a: 93, 100). Proops was also the first to notice Frege’s commitment to this view in *Begriffsschrift*.

³³ The attempting is critical. One will *in fact* refer to the True even if one utters “5 is prime” as the antecedent of a conditional.

(SM, op. 34). This, we think, is a promising and underappreciated idea, quite at odds with the way philosophers nowadays tend to think about truth. But we could hardly develop it here, even if we knew how to do so.³⁴

The Truth-Value Thesis is thus not just a semantical or logical doctrine. It is Frege's attempt to explicate the important but maddeningly difficult idea that judgment (belief, assertion) constitutively aims at the truth.

7.6 CLOSING: FREGE AND DEFLATIONISM

In closing, we want to say a few words about another issue related to the regress argument, namely, whether Frege was a deflationist. The question here is not whether the regress argument itself has such a conclusion. As we have said, we reject such interpretations. But some, such as Horwich (1990: 39), have a much simpler reason for labeling Frege a deflationist, namely, that he claims that the sentence "The thought that five is prime is true" has the very same sense as "Five is prime" itself (SM, op. 34).³⁵

The overall tenor of our discussion should have made it clear both that and why Frege is not a deflationist. Perhaps Frege did think that the predicate "is true" of natural language was redundant when applied to a thought explicitly identified by a clause expressing it.³⁶ But even so, his view, as we saw in the last section, was emphatically that this is not truth as logic knows it. For logic, the True is one of the two truth-values, and the truth-values are the referents of sentences. Logic would therefore be concerned with the True even if there were no truth-predicate as, indeed, there is not in Frege's logic.³⁷

Moreover, Frege clearly would not agree with the common deflationist thesis that there is nothing substantial to be said about what it is for a thought to be true. For Frege, the thought that five is prime has a structure corresponding to that of the sentence "Five is prime" (CT, op. 36). Roughly, the thought that five is prime is composed of the sense of "five" and the sense of the predicate "is prime." These, in turn, determine references: the number five and the concept *is prime*, respectively. These then compose

³⁴ Textor (2010) and Jarvis (2012) develop related ideas.

³⁵ See also development of this argument by Kemp (1998) and the reply by Heck (2002).

³⁶ The depth of Frege's commitment to this view is not clear, however. There are several places in Frege's writings where he identifies the senses of sentences when one might have thought he should regard them as distinct (Heck and May 2010: section 5). Our view is that Frege does so because he tends to conflate a sufficient condition for difference of sense—that it should be possible to believe one but not the other—with a necessary condition—that it should, in some more practical sense, be possible to believe one but not the other. Frege gives no argument for the necessary condition, which has no real plausibility and plays no significant role in his philosophy. It is the sufficient condition that does the actual work.

³⁷ The horizontal is not a truth-predicate, any more than " $\zeta = \forall x(x = x)$ " is.

via function-application, so as to determine that the thought that five is prime has as its referent the True (Heck and May 2010). This is a very long way from the view that “[t]he entire conceptual and theoretical role of truth may be explained” in terms of the assumption that it is true that five is prime if, and only if, five is prime (Horwich 1990: 39). It is, rather, the very birth of semantics.

REFERENCES

- Antonelli, A. and May, R. (2000). “Frege’s New Science,” *Notre Dame Journal of Formal Logic* 41: 242–70.
- Beaney, M. (2007). “Frege’s Use of Function-Argument Analysis and his Introduction of Truth-Values as Objects,” *Grazer Philosophische Studien* 75: 93–123.
- Boole, G. (1854). *The Laws of Thought*. London: Walton and Maberly.
- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Burge, T. (2005). “Frege on Truth,” in *Truth, Thought, Reason: Essays on Frege*. New York: Oxford University Press, 83–132.
- Davidson, D. (1984). “Communication and Convention,” in *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 265–80.
- Dummett, M. (1981). *Frege: Philosophy of Language*, 2nd edn. Cambridge, MA: Harvard University Press.
- Frege, G. (1972). *Conceptual Notation and Related Articles*, ed. and trans. T. W. Bynum. New York: Oxford University Press.
- Frege, G. (1979). *Posthumous Writings*, ed. H. Hermes, F. Kambartel, and F. Kaulbach, trans. P. Long and R. White. Chicago: University of Chicago Press.
- Frege, G. (1984). *Collected Papers on Mathematics, Logic, and Philosophy*, ed. B. F. McGuinness. Oxford: Basil Blackwell.
- Frege, G. (1997). *The Frege Reader*, ed. M. Beaney. Oxford: Blackwell.
- Frege, G. (AimCN). “On the Aim of the Conceptual Notation,” trans. T. W. Bynum, in Frege (1972), 90–100.
- Frege, G. (Bg). “*Begriffsschrift*: A Formula Language Modeled upon that of Arithmetic, for Pure Thought,” tr. S. Bauer-Mengelberg, in J. van Heijenoort (ed.), *From Frege to Gödel: A Sourcebook in Mathematical Logic* (1967), Cambridge, MA: Harvard University Press, 5–82.
- Frege, G. (BLC). “Boole’s Logical Calculus and the Concept-script,” trans. P. Long and R. White, in Frege (1979), 9–46.
- Frege, G. (BLF). “Boole’s Logical Formula-language and my Concept-script,” trans. P. Long and R. White, in Frege (1979), 47–52.
- Frege, G. (CSM). “Comments on Sense and Meaning,” trans. P. Long and R. White, in Frege (1979), 118–25. Also in Frege (1997), 172–80.
- Frege, G. (CT). “Compound Thoughts,” trans. P. Geach and R. H. Stoothoff, in Frege (1984), 390–406.
- Frege, G. (FC). “Function and Concept,” trans. P. Geach, in Frege (1984), 137–56. Also in Frege (1997), 130–48.
- Frege, G. (Ge01). “On the Foundations of Geometry: First Series,” trans. E.-H. W. Kluge, in Frege (1984), 273–84.

- Frege, G. (Geo2). "On the Foundations of Geometry: Second Series," trans. E.-H. W. Kluge, in Frege (1984), 293–340.
- Frege, G. (Gg). *The Basic Laws of Arithmetic*, trans. P. A. Ebert and M. Rossberg (2013). Oxford: Oxford University Press.
- Frege, G. (Gl). *The Foundations of Arithmetic*, 2nd revised edn, trans. J. L. Austin (1980). Evanston IL: Northwestern University Press.
- Frege, G. (Log97). "Logic (1897)," trans. P. Long and R. White, in Frege (1979), 126–51. Extracts in Frege (1997), 227–50.
- Frege, G. (PMC). *Philosophical and Mathematical Correspondence*, ed. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart, trans. H. Kaal (1980). Chicago, University of Chicago Press.
- Frege, G. (SM). "On Sense and Meaning," trans. M. Black, in Frege (1984), 157–77. Also in Frege 1997: 151–71.
- Frege, G. (Tht). "Thoughts," trans. P. Geach and R. H. Stoothoff, in Frege (1984), 351–72. Also in Frege (1997), 325–45.
- Frege, G. (WMR). "What May I Regard as the Result of My Work?," trans. P. Long and R. White, in Frege (1979), 184.
- Geach, P. T. (1965). "Assertion," *The Philosophical Review* 74: 449–65.
- Gödel, K. (1930). "Die Vollständigkeit der Axiome des logischen Funktionenkalküls," *Monatshefte für Mathematik und Physik* 37: 349–60. Translated as Gödel (1986b).
- Gödel, K. (1931). "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme, I," *Monatshefte für Mathematik und Physik* 38: 173–98. Translated as Gödel (1986c).
- Gödel, K. (1986a). *Collected Works*, 3rd edn, ed. S. Feferman, J. Dawson, and S. Kleene, vol. 1. Oxford: Oxford University Press.
- Gödel, K. (1986b). "The Completeness of the Axioms of the Functional Calculus of Logic," in Gödel (1986a), 102–23.
- Gödel, K. (1986c). "On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems I," in Gödel (1986a), 144–95.
- Heck, R. G. (1998). "Grundgesetze der Arithmetik I §§29–32," *Notre Dame Journal of Formal Logic* 38: 437–74.
- Heck, R. G. (1999). "Grundgesetze der Arithmetik I §10," *Philosophia Mathematica* 7: 258–92.
- Heck, R. G. (2002). "Meaning and Truth-conditions: A Reply to Kemp," *Philosophical Quarterly* 52: 82–7.
- Heck, R. G. (2010). "Frege and Semantics," in Potter and Ricketts (2010), 342–78. Originally published in *Grazer Philosophische Studien* 75: 27–63.
- Heck, R. G. (2011). "The Julius Caesar Objection," in R. G. Heck, *Frege's Theorem*. Oxford: Clarendon Press, 127–55.
- Heck, R. G. (2012). *Reading Frege's Grundgesetze*. Oxford: Clarendon Press.
- Heck, R. G. (2015). "Formal Arithmetic Before *Grundgesetze*," in P. Ebert and M. Rossberg (eds.), *A Companion to Frege's Grundgesetze*. Oxford: Oxford University Press (forthcoming).
- Heck, R. G. and R. May (2006). "Frege's Contribution to Philosophy of Language," in E. Lepore and B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*. Oxford: Oxford University Press, 3–39.
- Heck, R. G. and R. May (2010). "The Composition of Thoughts," *Noûs* 45: 126–66.
- Heck, R. G. and R. May (2013). "The Function is Unsaturated," in M. Beaney (ed.), *The Oxford Handbook of the History of Analytic Philosophy*. Oxford: Oxford University Press, 825–50.

- Horwich, P. (1990). *Truth*. Oxford: Blackwell.
- Jarvis, B. (2012). "The Dual Aspects Theory of Truth," *Canadian Journal of Philosophy* 42: 209–33.
- Kemp, G. (1995). "Truth in Frege's 'Law of Truth,'" *Synthese* 105: 31–51.
- Kemp, G. (1998). "Meaning and Truth-conditions," *Philosophical Quarterly* 48: 483–93.
- Kneale, W. and M. Kneale (1962). *The Development of Logic*. Oxford: Clarendon Press.
- Kremer, M. (2010). "Sense and Meaning: The Origins and Development of the Distinction," in Potter and Ricketts (2010), 220–92.
- Kripke, S. (1975). "Outline of a Theory of Truth," *Journal of Philosophy* 72: 690–716.
- Larson, R. and G. Segal (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.
- MacColl, H. (1877). "The Calculus of Equivalent Statements and Integration Limits," *Proceedings of the London Mathematical Society* 9: 9–20.
- May, R. (2001). "Frege on Identity Statements," in C. Cecchetto, et al. (eds.), *Semantic Interfaces: Reference, Anaphora, and Aspect*. Stanford, CA: CSLI Publications, 1–62.
- May, R. (2006). "The Invariance of Sense," *Journal of Philosophy* 103: 111–44.
- Post, E. (1921). "Introduction to a General Theory of Elementary Propositions," *The American Journal of Mathematics* 43: 163–85.
- Potter, M. and T. Ricketts (eds.) (2010). *The Cambridge Companion to Frege*. Cambridge: Cambridge University Press.
- Proops, I. (1997). "The Early Wittgenstein on Logical Assertion," *Philosophical Topics* 25: 121–44.
- Ricketts, T. (1986). "Objectivity and Objecthood: Frege's Metaphysics of Judgement," in L. Haaparanta and J. Hintikka (eds.), *Frege Synthesized: Essays on the Philosophical and Foundational Work of Gottlob Frege*. Dordrecht: Reidel, 65–95.
- Russell, B. (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- Schröder, E. (1972). "Review of Frege's *Conceptual Notation*," trans. T. W. Bynum, in Frege (1972), 218–32.
- Stanley, J. (1996). "Truth and Metatheory in Frege," *Pacific Philosophical Quarterly* 77: 45–70.
- Sullivan, P. M. (2005). "Metaperspectives and Internalism in Frege," in M. Beaney and E. H. Reck (eds.), *Gottlob Frege: Critical Assessments of Leading Philosophers*, vol. 2. London: Routledge, 85–105.
- Sundholm, G. (2001). "Frege, August Bebel, and the Return of Alsace-Lorraine: The Dating of the Distinction between *Sinn* and *Bedeutung*," *History and Philosophy of Logic* 22: 57–73.
- Tappenden, J. (1997). "Metatheory and Mathematical Practice in Frege," *Philosophical Topics* 25: 213–64.
- Tarski, A. (1958). "The Concept of Truth in Formalized Languages," in J. Corcoran (ed.), *Logic, Semantics, and Metamathematics*. Indianapolis: Hackett, 152–278.
- Textor, M. (2010). "Frege on Judging as Acknowledging the Truth," *Mind* 119: 615–55.
- Venn, J. (1972). "Review of Frege's *Conceptual Notation*," trans. T. W. Bynum, in Frege (1972), 234–5.
- Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness. London: Routledge and Kegan Paul.
- Wittgenstein, L. (1979a). *Notebooks: 1914–1916*, 2nd edn, ed. G. H. von Wright and G. E. M. Anscombe, trans. G. E. M. Anscombe. Chicago: University of Chicago Press.
- Wittgenstein, L. (1979b). *Wittgenstein's Lectures: Cambridge 1932–1935*, ed. A. Ambrose. Chicago: University of Chicago Press.

PART III

.....

THE CLASSICAL
THEORIES OF TRUTH

.....

CHAPTER 8

THE COHERENCE THEORY OF TRUTH

RALPH C. S. WALKER

A coherence theory of truth is a theory that holds truth to consist in coherence. To understand the nature of truth is to understand that it is coherence, or a certain kind of coherence, very much in the way that to understand the nature of water is to know that it is H_2O . The only difference is that the identity of water with H_2O was discovered scientifically, whereas in this case the identity must be established by philosophical means.

The coherence theory has often been felt to be counterintuitive, and has sometimes been briskly dismissed for that reason. Yet it keeps reappearing in new and challenging forms. Part of the reason for this is that the alternative theories have their difficulties too, but there are also powerful philosophical pressures that have pushed philosophers of very different backgrounds to adopt a coherence theory of truth. The nineteenth-century idealists found themselves driven that way; a little later the verificationists found themselves under very similar pressures, and Neurath in particular recognized that these pressures must inevitably lead them to a form of coherence theory. More recently these same pressures, in slightly different forms, have led Putnam and Davidson explicitly to adopt coherence theories of truth, and have caused anti-realists like Michael Dummett to adopt what amounts to a coherence theory in all but name. These pressures derive from two closely related thoughts, one epistemological and the other semantic. The first is the thought that there is something quite unsatisfactory, even absurd, about the skepticism that suggests our most carefully adjusted beliefs about the world could actually be false. The second supports this by considerations about meaning: the meaning of our words must be tied to experience in a way that precludes us from any such radical error.

Coherence theorists have always found difficulty with the correspondence theory of truth, which they see as their real rival. The correspondence theory has trouble explaining what “facts” are and what it is for a fact and a truth to “correspond.” Coherence theorists believe these difficulties arise from a mistaken conception of reality. This is not to say that coherence theorists repudiate such obvious truisms as “true statements correspond with the facts,” when these are meant just as truisms. What they

repudiate is any attempt to turn such truisms into serious theories. So long as by “facts” we just mean what true statements state, and have nothing more substantive in mind by “correspondence,” such truisms are harmless but trivial—they do not constitute any theory at all.

The correspondence theory goes beyond the truisms by its account of what correspondence is, and of what is supposed to correspond with what. The coherence theory likewise seeks to give an account of coherence, and of the items that are supposed to exhibit this coherence. Both theories share the assumption that truth *has* a nature, and that problems about truth have a metaphysical character: they concern the relationship between what we say or think and the reality that we are seeking to describe. Both theories hold that there is more to be said about this relationship than can be captured by Ramsey’s redundancy theory, or any of the theories following it that have sought to dissolve the real problem of the nature of truth by looking at how we use expressions like “is true.”

Being a theory of the *nature* of truth, the coherence theory must be sharply distinguished from the theory that coherence is often, or always, a *guide* to the truth. Clearly we often rely on coherence as a guide to the truth, but that commits us to no particular view of what truth is; only to the belief, or the hope, that coherence and truth are linked in some fairly reliable way.¹ The coherence theory of truth goes beyond that, and claims that truth just *is* coherence, or coherence of some specific kind.

There is no single coherence theory of truth, just as there is no single correspondence theory. Different accounts of correspondence have been put forward; so have different accounts of coherence, though the different accounts do have a certain amount in common. Coherence requires consistency, and accordance with the laws of logic—though there is certainly some disagreement as to what the laws of logic are, the law of excluded middle being particularly open to question. But further principles of coherence are needed. Most would find them in the fundamental principles that govern our everyday and scientific thinking; there is rather more room for dispute as to just what these are.² What is shared by all coherence theories of truth is the view that the principles of coherence make no claim to match some mind-independent reality. They articulate what truth is, and they are themselves constitutive of truth. Truth is thus determined by these standards, which are the standards of rationality; the standards which govern our thought, insofar as we can claim to be rational. This is where the contrast with the correspondence theory lies. For the correspondence theory, truth consists in correspondence with a reality that exists independently of our thought about it.

¹ In his (1939), Blanshard argued that a coherence theory of justification unavoidably leads to a coherence theory of truth. His ground was the epistemological one, that our tests for truth would not ensure that truth was attained unless truth consisted in coherence.

² Detailed work is of course needed on how to develop these principles into a satisfactory standard of coherence. Cf. Olsson (2002).

What sort of items do coherence theorists take to cohere? They have sometimes said that truth consists in the maximally coherent system of propositions. This however will not do, at least if propositions are supposed to be abstract entities and not things that are actually believed or asserted by anyone. As Russell pointed out, the (false) proposition that Bishop Stubbs was hanged belongs to as large a coherent set of propositions as the (true) proposition that Bishop Stubbs died in bed.³ These maximally coherent systems of propositions would describe alternative possible worlds.

Joachim (whom Russell was attacking) conceived of the truth not as a maximally coherent system of abstract propositions, but as something like a maximally coherent system of judgments—by which he meant, judgments that people were prepared to make.⁴ Sometimes the coherence theory has been expressed in terms of beliefs, which comes to much the same. The idea needed is just that of what people accept to be the case, where (roughly at least) someone who accepts something is prepared sincerely to assert it, or to assent to it sincerely when it is asserted. I shall use the terms “judgment” and “belief,” because other writers do, but I shall also use the word “acceptance,” treating these three terms to be equivalent. The more neutral word “acceptance” may help to avoid any misleading connotations of “belief” and “judgment.”

Coherence theories of *truth* need to be distinguished from coherence theories of *reality*. Coherence theories of reality are not theories about the nature of truth; they are metaphysical theories to the effect that reality itself meets certain standards of coherence. Sometimes they are theologically grounded, the creator being thought to have devised a rationally coherent universe; but they need not be. The metaphysical claim is just that the standards of rational coherence reveal to us reality as it is, because it is in the nature of reality to exhibit just this coherence.

Coherence theories of truth and coherence theories of reality can look very similar: indeed they are often confused. Spinoza, I believe, held a coherence theory of truth, but others think he only held a coherence theory of reality. Bradley held a coherence theory of reality, not (despite what is often said) a coherence theory of truth: he thought that reality itself, quite independently of us and of our ways of thinking about it, meets standards of coherence that render it comprehensible to us.⁵ A coherence theory of reality leaves open the question why our standards of rational coherence should match the way the world is, though the holder of it may be able to provide a metaphysical answer (perhaps God has made us think in ways that ensure our thoughts will match that independent reality). A coherence theory of truth leaves no room for the question to arise. Truth just is coherence—a coherence that we can recognize because our own standards determine it. Hence if we have a coherent system there is no further question whether

³ Russell 7.

⁴ Joachim (1906).

⁵ Walker (1998). I have discussed Spinoza in Walker (1989: ch. 3).

it is true: it has to be, since coherence is the nature of truth. For the coherence theory of reality, certainly, if we have a coherent theory we can be assured it is true, but that is because we are confident that reality is coherent. There is still room for the question whether reality really *is* coherent, or coherent by the standards we are using.

Thus the coherence theory of truth dissolves an epistemological problem, which the coherence theory of reality leaves in place. The problem is Descartes's problem, of how those things "which I think I see utterly clearly with my mind's eye"⁶—the things which are evident to him by the natural light of reason—can be guaranteed to be true; or in other words, how the standard we use can be guaranteed to give us truth about the way the world is. He tries to resolve it by proving the existence of a God who is no deceiver. But that proof has often been thought circular, and Descartes's solution has often seemed unsatisfactory. It seemed unsatisfactory to Kant.

Kant took the problem very seriously, but his ultimate reaction was to dissolve it, by saying that truth, truth as we ordinarily conceive it, just *is* that system of beliefs that meets with a certain standard of coherence that is determined by the way our minds operate. It has to be, because the world as we can know it must be reality as conditioned by human cognitive powers; the conceptual ordering of our thought, and the ordering of our experience within a spatio-temporal framework, are features of *ourselves*, rooted in the nature of our capacities. They cannot be assessed as correct or incorrect beyond saying that they are inevitable for us. And there is no other way that we could know the world, for we can know the world only in the ways that our cognitive capacities provide. Hence the world as we can know it, the everyday world of people and galaxies and bacteria, just *is* the world so ordered and conceptualized. This *removes* the problem of how to justify supposing that the ways in which we think and register our experiences really do match reality. The truth about the world is constituted by that systematically ordered pattern of things that our minds can grasp, taking account of sensation but finding regular connections within our experience in accordance with our minds' own ways of working.

This amounts to a coherence theory of truth. But it is not a coherence theory of truth in a quite unqualified form. The coherence required is not just with "the formal conditions of experience," which have to do with human cognitive capacities, but also with "the material conditions of experience, i.e. with sensation" (A 218/B 265–6).⁷ It is important that sensation is a given element in experience. Sensation is not constituted by coherence. Moreover, for Kant there is still a reality that is radically independent of our ways of thought, and which he takes to be the "ground" of the sensations we receive. Only, about that radically independent reality, that reality of "things as they are in themselves," we can know nothing at all. This need not bother us, in his view, because truth in the only sense in which it makes sense to worry about it *is* just what the coherent operation of our principles yields.

⁶ Descartes, (1964–76), VII:36.

⁷ Kant (1929), trans. Kemp Smith.

Kant's idealist successors, by and large, thought this admission of things in themselves was a mistake, and that led them to a quite unqualified coherence theory of truth. *Because* truth consists in coherence, reality—the only reality there is or could be—is just what a thoroughly coherent system of beliefs would describe. The idea of things in themselves must then be incoherent; to admit them would be to indulge in empty play with words. Kant had held that we can *think* intelligibly about things as they are in themselves, but the idealists rejected this. To suppose that there is a realm of Kantian things in themselves is to fail to suppose anything, for we have no intellectual resources to describe it: even to say it contained “things” would be to transpose a familiar word to an entirely alien setting in which no rules for its use remain.

It is with the nineteenth-century idealists, and with their spiritual forebear Spinoza, that it is often hardest to draw the line sharply between a coherence theory of truth and a coherence theory of reality. This is because for many of them, including Spinoza, the rational coherent system that constitutes the truth is not (as it was for Kant) the product of *our* minds, in the sense that its principles are determined by our capacities and our ways of thinking. Instead it is the product of the mind of God (or Spirit, or the Absolute), who is far more rational than we are.

If we left it at that, we should not really have a coherence theory of truth, any more than we have a coherence theory of truth if we hold that a rational God has designed the universe in a rationally intelligible way. That would be a coherence theory of reality, not of truth, because it still leaves open the question of whether our minds are capable of grasping that rational order of things—the question of whether our rationality and God's rationality are the same. Descartes held a coherence theory of reality, or something close to it: he thought the world was intelligibly constructed by a wholly rational God. He also thought that rational insight gave us the basic principles of geometry and mechanics, and thus a fundamental description of the physical world. Yet for Descartes this insight was dependent on God's being no deceiver. It was insight about a reality altogether independent of us.

For Spinoza, however, and for many of the later idealists, there is no need for an assurance that God is no deceiver. The potential failure of match between our ways of thinking and God's ways of thinking—our standards of coherence and his—does not and cannot exist. For our minds are parts of God's mind, finite centers of awareness within an infinite whole. Spinoza maintains that to the extent that our ideas are “adequate”—wholly clear and distinct—they just *are* God's ideas, which are in turn to be identified with the physical or mental reality that they describe. So “truth is its own standard”;⁸ it is not a matter of one thing matching another. Yet Spinoza is not unequivocal about this. That is partly because he had not thought his position fully through, but partly also because of a problem which remained serious for the later idealists, and which led Joachim reluctantly to conclude that, after all, no coherence theory of truth can ever be quite satisfactory. We have many ideas that are far from “adequate”: we make

⁸ Spinoza (1985), ed. Curley, II Prop. 43.

many judgments that are false. Spinoza runs into serious difficulty over false judgments, which he has to regard as “incomplete,” as well as over how we can tell which of our ideas are adequate. For Joachim (and indeed for common sense) there always remains something about our judgments that is ineluctably “other” than that which our judgments are about. It is because of this “otherness” that our judgments can fail to be true, and it is because of this “otherness” that we can never be confident that our internal ideas of coherence are at all adequate to the reality we seek to describe. Joachim ends his book by reiterating his conviction that “the truth itself is one, and whole, and complete,” in other words, wholly coherent. But that is really an affirmation of a coherence theory of reality. “[N]o theory of truth as coherence can be completely true; for as a system of judgments, as a piece of discursive knowledge, it must be ‘other’ than the truth ‘about’ which it is, and thus it must fail of that concrete coherence which is complete truth”—or as we might prefer to say, “which is complete reality.”⁹

Joachim did want to hold a coherence theory of truth, though worried by his own objection. In fact his objection does not really refute the coherence theory of truth. What it shows is that a coherence theory of reality that leaves reality independent of our thought gives us no reason to suppose we can find truth by relying on our own standards of coherence. The standards of rational coherence that we find compelling in ordering our thoughts may be very imperfectly aligned with the standards of rational coherence that are reflected in the universe, or in the mind of God. It does not help to say that our minds are in some sense part of God’s mind; there remains an epistemological gap between our judgments and the way things are.

Since the coherence theory of truth holds that it is the principles of our thinking that determine the truth, it is not surprising to find it associated with philosophers who describe themselves as idealists. But it has also attracted others who would have found that label uncongenial. To reject the coherence theory of truth is to hold that in talking about the world we are making claims about a reality that is wholly independent of what we think or say about it. We are perfectly happy to accept that reality is independent of any particular judgments that we make, but it does seem a lot less satisfactory to have to accept that our best, our most “coherent,” theory about the world could turn out to be just wrong, even where that theory has been framed and tested under the best possible conditions. This seemed unsatisfactory to the idealists, primarily on epistemological grounds: it leaves room for a kind of skepticism that seems absurd, the kind that suggests we could be wrong about everything. Others go further, and base their rejection of such skepticism on semantic considerations. In their view, to think there is room for it is to ignore the preconditions for language use. Words can only acquire meaning through being used in publicly observable circumstances, and these publicly observable circumstances must provide them with their meaning. They cannot acquire meanings that could not be learned and manifested in this way, so they cannot (it would seem)

⁹ Joachim (1906: 178).

have meanings that transcend what can somehow be cashed out in terms of our shared experience. The idea of a reality wholly independent of what we can find out about does not make sense.

Such considerations motivated Putnam's adoption of a coherence theory of truth in the 1980s (though he gave it up later).¹⁰ The idea of a reality metaphysically independent of us, and about which our best theory might be quite mistaken, seemed to him untenable in its own right, and incompatible with a workable account of how we can learn and use language. He questioned how we could ever use our language to refer to something in that independent reality. We learn language through using it in familiar kinds of recognizable circumstance; how could the word "cow" acquire a reference to something in a mysterious independent reality? In a similar way, many philosophers have felt that there is "no vantage point outside our conceptual scheme." Davidson explicitly adopted a coherence theory on grounds partly similar to Putnam's, in that coherence seemed to him the only basis for an understanding of language, of one another's beliefs, and of reality.¹¹

Putnam and Davidson—and sometimes Quine as well—share with Joachim and with many others the idea that *all* truth consists in coherence. This at least cannot be right. A coherence theory of truth which is intended to apply globally—i.e. to truths of all kinds without exception—must itself be incoherent.

A coherence theory of truth that extends to *all* truths cannot be sustained, because there must be some truths that determine the character of the coherent theory itself, and *their* truth therefore cannot be determined by it. The coherent system of truths must be determined by people's acceptances—the propositions or statements that they actually accept—and by the principles and rules of inference that they use to determine what is to constitute coherence within the system. There must therefore be truths *about* what it is that people accept, and these truths about acceptances cannot consist in coherence. A coherent system of purely abstract propositions may include propositions like "everyone accepts that pencils are made of cheese," "everyone accepts that everyone accepts that pencils are made of cheese," and so on. What we need is not what the system itself declares to be beliefs, acceptances, or rules, but *actual* beliefs, *actual* acceptances, rules that are *actually* treated as in force. *That they are actually held* is something the system must take as a given. The system can then determine truths in other cases. But it cannot determine the truths on which it rests, the truths about what is actually accepted and about what rules are actually treated as norms. To these truths the coherence theory cannot apply. They are true *because* that is the way the world is.

Attempting to make the coherence theory global—to make it apply to all truths whatever—would lead to an incoherent regress. Let us suppose that P is accepted. If the

¹⁰ Putnam (1981); (1983). For his repudiation of the position, Putnam (2001).

¹¹ Davidson (1984); (1986).

truth that P is accepted consists in the coherence of “P is accepted” with the acceptance of Q, R, S, and principle of inference R_1 , then we are treating it as true that Q, R, S, and R_1 are themselves accepted. But if the acceptance of Q consists in its coherence with the acceptance of P, R, S, R_1 , “P is accepted,” “It is accepted that P is accepted,” and so on, we have a circle of mutually supporting propositions. “Q is accepted,” “R is accepted,” “S is accepted,” “ R_1 is accepted,” “It is accepted that it is accepted that it is accepted that T is accepted,” and so on, will belong to the same circle. But we never get down to the *fact* that any proposition or rule is *actually accepted*. We are back with Russell’s original objection about Bishop Stubbs. There can be indefinitely many collections of mutually supporting propositions each describing a coherent possible world, and only one of them describes the actual world. We sought to avoid that by saying the coherence theory should not be expressed in terms of abstract propositions, but rather in terms of beliefs or acceptances—propositions that real people actually do accept. It is the fact that real people actually do accept P, Q, R, S, and R_1 that makes it possible for someone to use them to construct a coherence theory of truth. And that fact cannot itself consist in coherence. If it did, then we could equally have a rule R_2 such that not-P, not-Q, not-R, not-S, and R_2 formed a coherent set, to which equally belonged “not-P is accepted,” “not-Q is accepted,” “It is accepted that it is accepted that not-S is accepted,” and so on.

James Young suggests getting round this by sharply divorcing *facts* from *truths*.¹² There could then be facts about acceptances, and the coherence amongst them could determine the truths, including whatever truths there may be about acceptances. But this makes the relationship between facts and truths mysterious, and deprives the coherence theory of its point. In its global form, the coherence theory seeks to do away with questions of correspondence between facts and statements generally; in a less than global form, it seeks to do away with questions of correspondence between facts and statements of a particular class—statements about the physical world, perhaps, or about mathematics. Young seeks to defend the global form. But if his suggestion were right, questions of correspondence could still be raised, and all he is really saying is that we should not call them questions of *truth*. We should reserve the word “truth” for coherence-truth.

That proposal seems very odd, but it could only constitute a verbal point at best. Young recognizes that facts about acceptances are needed for the theory to get going, and that there are statements that describe these facts. His theory has to allow that there is room for the correspondence relation here; coherence-truth alone will not do. To reserve the word “truth” for coherence-truth alone does not help, it just adds confusion. The coherence theorist should admit that there are facts, and therefore truths, about acceptances to which his theory does not apply.

A common objection to global coherence theories—made e.g. by Schlick—is that they cannot give an appropriate place to observation.¹³ For they do not allow us to

¹² Young (2001).

¹³ Schlick (1934); trans. in Ayer (ed.) (1959).

treat it as a basic *given* that a certain observation occurs: to treat it in that way would be to allow it a factual status that did not consist in coherence. But once we see that a coherence theory cannot really be global in any case, because there must be truths about acceptances which cannot themselves consist in coherence, this no longer looks like an additional objection. It has been provided for. The acceptances that have to be treated as given will naturally include a large number of perceptual beliefs, at various levels of sophistication. I believe that there is a sheet of paper in front of me; I would also assent to the proposition that it seems to me that I see a white rectangular shape. The datum is not of course that I really do see anything of the sort, or that there really is a sheet of paper in front of me; what is to be treated as given is just that these are amongst my acceptances. But if we agree to that, we allow experience to play its proper part in framing the truths about the world, without any further concession by the coherence theorist than was already required by the admission of a special status for acceptances.

Schlick was arguing against Neurath. Both were verificationists, but Neurath held that verificationism committed one to a coherence theory of truth—and one which made *all* truths consist in coherence.¹⁴ Neurath was partly right: verificationism does require a coherence theory of truth. But not one that makes all truths consist in coherence. Schlick was right about the special status of observation.

Schlick thought that our understanding both of language and of the world must be based on observational reports that could not be mistaken: statements like “Here yellow borders on blue” and “Here now pain” are guaranteed of truth when uttered sincerely. Others have thought something similar. Descartes’s *cogitationes* played a foundational role for him, because he thought they allowed for no possibility of error. Neurath argued that there are no such foundational beliefs. Both he and Schlick took the issue between them, over the coherence theory of truth, to depend upon this. It does not.

Descartes wanted incorrigible beliefs in order to provide the foundation for knowledge. Few would now place such reliance on incorrigible beliefs, but however that may be, there is a difference between a theory of knowledge and a theory of truth. Incorrigible beliefs are not required as the foundation of a coherence theory of truth. What is required is that a special status be given to acceptances, including observational acceptances: for claims about what is accepted, truth must consist in correspondence and not coherence. But no such claims need be incorrigible. In assessing claims about what is accepted we normally rely on their coherence with the rest of our beliefs (have we any reason to think the speaker is joking, or trying to mislead us about what she thinks?) and on our principles of inference, but it is not these that make the statement true or false: they only contribute to our *grounds* for believing it true or false. Its truth or falsity consists in its correspondence, or its failure to correspond, with an independent reality.

¹⁴ Neurath (1932–33); trans. in Ayer (ed.) (1959). Carnap and Hempel came to a similar conclusion.

It is *truths about acceptances* that have this special status. Schlick was interested not in claims about who has what observational beliefs, but in the observational beliefs themselves. It was those he thought had to be incorrigibly true. They do not; nor does their truth have to consist in correspondence with fact. But for a verificationist, and perhaps for anyone who takes experience seriously, they have a different sort of foundational role, one that again does not require incorrigibility. They provide the foundation (or a substantial part of the foundation) for our understanding of language. Verificationism holds that thought can be expressed only in language, and that the way in to understanding language must be by understanding the meaning of those sentences that are used to make assertions.¹⁵ To understand the meaning of an assertoric sentence is to know how to recognize the conditions that verify it or falsify it on particular occasions of utterance. There are some sentences which on appropriate occasions can be used to make “observation statements,” expressing observational beliefs—on these occasions the sentences are uttered in circumstances in which their meaning can be taught and learned directly, by their association with a particular type of experience. This does not mean that these observation statements are true every time. “That is a sparrow” could be used, on suitable occasions, to teach the word “sparrow,” but there will be occasions on which the speaker gets it wrong, having misidentified the bird herself, or mistaking a plastic sparrow for a real one. We manage to teach the language despite such cases. To know how to recognize the conditions that verify or falsify a claim is rarely if ever to know how to verify it or falsify it decisively by a single observation; it is to know, among other things, what circumstances can turn up to defeat it—what further investigations might show that it was not a sparrow after all. But it seems evident enough that some sentences, a subset of those that describe things we can directly observe, must provide a basis for our understanding of language.

There are of course very many sentences which, whenever and wherever uttered, make claims with a content that goes well beyond the possibility of simple observation. We use such sentences for many of the statements we make about everyday life, and most of the statements we make in science. Verificationists were committed to holding that the meaning of any such sentence must still be identified with the method of verifying or falsifying it—or determining whether its assertion would be warranted or not, since conclusive verification would commonly be impossible. Since this requires going beyond observation, there must be rules relating observations (or observation statements) to the verification, or confirmation, of claims of this kind. For any such claim the relationship to observation may be rather remote, though being firm empiricists the verificationists were confident that there must be one. It could be mediated by a complex network of non-observational claims, provided that they in turn derived their meaning ultimately from a relationship to experience.

¹⁵ As I am using the terms, a “sentence” is a series of words appropriate for expressing a complete thought; the same sentence can be used on many different occasions. If it is used assertorically a sentence may be used on a particular occasion to make a “statement,” and it is statements—tied down to particular contexts and occasions of utterance—that are capable of having truth-values.

What sort of rules could relate these claims together? This is the question we asked earlier about principles of inference, when we were distinguishing coherence theories of truth from coherence theories of reality. What makes them good principles of inference to choose? Coherence theories of *reality* take them to be good just in case they are correct. Whether or not they are correct then depends on whether they match a reality that is independent of us, and provide us with a reliable way of reaching truths that are themselves independent of us and of our ways of thinking. Coherence theories of *truth* take them to be good just because they reflect our own standards. They are thereby determinant of truth, truth thus being constituted by our standards of coherence.

Verificationists are committed to the second alternative. For they are committed to holding that what makes the rules correct is that they are built into the meaning of the non-observational sentences. If they were not correct, these sentences would have to be meaningless—we should have no method of verifying or falsifying them. To understand a non-observational sentence is to understand how it can be confirmed by observation, and this means understanding the rules that exhibit how observations can provide its confirmation.

Could the rules nevertheless correspond with an independent reality, in that it was a feature of that independent reality that these rules took us from truths to truths, or at any rate from likely truths to likely truths? For the verificationist the answer must be “No.” These rules themselves can be stated in sentences, and if those are to be capable of bearing truth-values their meanings must (for the verificationist) consist in the methods by which they can be verified. But there is no way in which they can be verified. If they could be verified, it would have to be either directly by observation, which is not possible since they express the rules that take us beyond observation; or else by a combination of observation and these rules themselves. That will not do because it would be circular. These are the rules by which we determine the meanings of non-observational sentences, and which therefore determine how *they* can be verified, but any demand for verification of the rules themselves is out of place. They are just rules.

But rules they are, and they are our rules. Our language depends upon them, and therefore our thought. By determining the meanings of non-observational claims, they determine what constitutes their truth. If there were any possibility that the rules were wrong, it would mean that our standards of verification were wrong. But our standards of verification determine the meanings of our words and thereby determine the content of our thoughts. There is no basis for assigning any kind of truth to these rules that is independent of our ways of thinking. The result is a coherence theory of truth. Not one, though, that makes *all* truth consist in coherence; the special status of truths about observational beliefs, and acceptances more generally, must remain.

And that must be a matter of serious concern for the coherence theorist, and not least for one whose coherence theory is founded on verificationistic considerations. We make statements about our own acceptances, and about those of other people. The same linguistic considerations should apply to them as apply elsewhere. So should truth for them not consist in coherence too? And yet it can't, or the coherence theory could not get going.

The verificationism of Schlick's and Neurath's time ran into difficulties because it worked with too tight a notion of meaningfulness, and had trouble giving a precise account of verifiability. Its commitment to a coherence theory of truth is however inherited by the form of anti-realism held to by Michael Dummett.¹⁶ Dummett starts from the thought that our understanding of the meanings of our words must be based on our ability to recognize when sentences containing them can warrantably be asserted. Warranted assertibility is a less demanding standard than verification: it may be very much less than conclusive. To understand a sentence one must have "an ability to recognize evidence for it when presented with it, and to judge correctly whether or not it is outweighed by any given piece of counter-evidence."¹⁷ But the standards of correctness are built into what it is to understand the sentences concerned. Like the verificationists, Dummett makes no claim to a correctness that is independent of our language and the ways of thinking that it encapsulates. If an assertion can by these standards be warranted conclusively, then there is no possibility, however slim, that it should fail to be true. Truth just *is* warranted assertibility, when the warrant determined by our standards is conclusive.

So Dummett's anti-realism is again a coherence theory of truth.¹⁸ It is opposed to realism, which he defines (for a given class of statements) as the belief that statements of the relevant class "possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us."¹⁹ For Dummett, on the contrary, unless there is discoverable evidence for or against a statement, that statement can have no truth-value. And what he means by that is what we should ordinarily mean: there is no such fact as the statement (or its negation) purports to describe. He is not just expressing ignorance as to whether the statement has a truth-value or not. If there is no evidence, there is no truth-value. The statement is neither true nor false. Physical reality, according to Dummett, contains "only what there is evidence that it contains."²⁰

It is not the individual who finds the evidence, or who decides what constitutes warrant, but the community of language-users, the community which assigns to a sentence a particular set of assertibility-conditions.²¹ I cannot have warrant for an assertion if other members of the same speech community would not equally have warrant for it under the same recognition conditions. Typically the evidence for and against the assertion needs to be weighed up, and the standards by which this weighing up is done

¹⁶ Crispin Wright's anti-realism needs to be distinguished sharply from Dummett's. It is not a coherence theory, so I shall not discuss it here. His article on Putnam, Wright (2000), is particularly illuminating in this context, however.

¹⁷ Dummett (2006: 59).

¹⁸ He denies this, but only because he takes it that the coherence theory must be expressed in terms of propositions, a view we have seen reason to reject.

¹⁹ Dummett (1978: 146).

²⁰ Dummett (2005: 679).

²¹ Dummett (2004: 67); (1991: 312f).

are again standards built into the language and therefore set by the speech community. These standards must be standards of coherence, for they are not answerable to anything beyond themselves, only to their own internal fit with one another. Thus my acceptances can be true, or warranted, only if they cohere with the standards set by the language, which themselves are standards of coherence.

Like the verificationists, Dummett seeks to provide what amounts to a global coherence theory. But he must be wrong about that, since a global coherence theory is not tenable. There must be truths about what is accepted, and no coherence account of them will do. Any statement I make about an acceptance, my own or somebody else's, may be corrigible by reference to further evidence—by looking to its fit within the system. But its truth cannot consist in that. The justification for holding it consists in its coherence with the system, but its truth does not. For this kind of statement there must be corresponding facts, or there cannot be a system of acceptances for other truths to cohere with.

But Dummett is not just interested in a global coherence theory. He is also interested in coherence theories of truth that apply within particular restricted domains, like those of logic and mathematics. I shall not comment on those further here, except to say that there is no obvious reason why truth in those fields, or perhaps in ethics and aesthetics, might not consist in coherence, even if truth about most other things does not. We do, after all, have some inclination to feel that there is a difference between the nature of mathematical truth and the nature of truth about rocks and stones and trees. To a greater or lesser extent we may feel that in other domains too. But Dummett certainly wants to provide an anti-realist account of rocks, stones, trees, and the physical world in general, and here he faces a difficulty that arose for the verificationists. For whatever we may think about mathematics, we do not have much inclination to think that there is a difference in kind between truths about people's acceptances and other truths about the world around us. Still, it is worth exploring the other difficulties that arise for Dummett's account of the physical world, since they will be difficulties for anyone who seeks to give a coherence theory of truth for the physical world.

The first of them is often raised, but it is not I think serious. It is the apparent threat of relativism. If warrant is determined by the community of language-users, might we not have a different truth for every different community of language-users? They could, it might seem, incorporate different standards of coherence into their use of language. Dummett would reject that, and plausibly. He regards the community of language-users as extending to cover all who use language at all. The intertranslatability of languages may offer some support for this, but there is more important support in the arguments adduced by Kant, and more recently by Strawson and others,²² to the effect that the principles constituting our standards of inference are conditions for anything that could count as intelligible thought at all. This is not the place to discuss these

²² Strawson (1966), and the considerable literature on transcendental arguments in recent years.

arguments, only to observe that it is far from obvious that a coherence theory of truth needs to be committed to allowing that there may be different coherent systems of truths for different groups of people.

There is a much more serious problem over what is to constitute the evidential base for the coherent theory. It is illustrated by the contrast between two short books of Dummett's. At one time he had been tempted to think that a consistent anti-realist should hold that statements about the past can have truth-values only on the basis of present evidence for them. In *Thought and Reality* he recognizes that would be "an intolerable position to adopt,"²³ because it so strongly conflicts with what we ordinarily think, and so he accepts that both *past and present* evidence is determinant of truth—but not future evidence since the future has not happened yet. Thus reality is "cumulative": "a proposition *becomes* true, and therefore comes to state a fact, only when evidence of its truth becomes available to us." However in *Truth and the Past* he repudiates this in favor of treating past and future symmetrically, allowing that "a proposition, whether about the past, the future or the present, is true, timelessly, just in case someone optimally placed in time and space could have, or could have had, compelling grounds for recognizing it as true—that is to say, if such compelling evidence would be or have been available to him."²⁴

A sizeable problem remains whichever alternative we choose. Besides the problem of going beyond what is *present*, we have the problem of going beyond what is *actual*. Dummett's talk of evidence may obscure the fact that the theory is grounded in the *recognition* of evidence—in people's observational acceptances. The actual observational acceptances that are made, by anyone, anywhere, and at any time, will still leave vast numbers of things indeterminate that we would normally take to be straightforward matters of fact; and this might be thought as "intolerable" as denying truth-values to past or future statements. They will be indeterminate because there is no evidence for or against them, so that there is nothing for their truth or their falsity to consist in. They do not cohere, and their negations do not cohere either, with the body of acceptances. I once slipped on a boulder near the top of a remote mountain, and the boulder rocked—it has two stable positions. But it is enormously unlikely that any actual evidence report could ever tell us which of the two positions it is in at present. It is most improbable that there is anyone anywhere near it. So there seems little chance that it can be determinately true or false that it is in the position in which I left it; or in the alternative position. This is not only strongly counterintuitive, but also conflicts with other principles that are built into our system for thinking about the physical world, for by induction we know that the position of boulders is always determinate.

Dummett seeks to deal with such cases by allowing not just actual but also *potential* evidence reports: those that *would* be made by "someone optimally placed in time and space." But what warrant could we have for assertions about what such hypothetical

²³ Dummett (2006: 74).

²⁴ Quotations are from the Preface to (2006, viii), where he contrasts the two approaches.

people would say? His answer is that we may have “indirect evidence,” for example of an inductive kind. “Any ground we have for supposing such-and-such a physical state of affairs to have obtained will also be a ground for supposing there to have been a possible observation in this sense, so that an observer located at a suitable space-time point could have observed that the state of affairs obtained.”²⁵ But this does not help at all with the boulder. For the problem is that we have perfectly good inductive grounds for supposing that the boulder is, now, in one or the other position, but no grounds for thinking that it is in this position rather than that.

The alternative is to say that there are really no grounds, and therefore no truth of the matter. This is what Dummett once said about the past; then about the future; now about neither. How far is common sense to guide us in such cases? It seems very remote from common sense to claim that there may be no truth of the matter in the case of the boulder.

This is not just a problem for Dummett. Any coherence theorist has to determine what acceptances—and whose acceptances—constitute the basis for the coherent theory. It arises for Kant, in that truth in the world of appearances extends only so far as “possible experience” extends, and Kant is very vague about what “possible experience” amounts to. Joachim was the first to raise the problem clearly, and it gave him further grounds for worry about the coherence theory. It does not help to say, as Putnam did, that truth is “an *idealization* of rational acceptability,”²⁶ where the standards are an idealized version of our present standards, internally corrected so as to need no further revision, and the evidence is such as would be needed for a fully developed science. That would require vastly more empirical evidence than we can reasonably suppose humanity will ever acquire. It represents, as Putnam said, an idealization, not a set of acceptances that are or ever will be actual. In fact it leaves us with the same problem as we had with those who equated the coherent system with the thought of God. The aim of the coherence theory was to understand truth in terms of *our* evidence and *our* standards, not in terms of a reality independent of us; and Putnam’s idealized science is as independent of us as is the mind of God.

There are two other problems for any coherence theory of physical truth that does not claim to be global, but allows that there are correspondence-truths about acceptances, including observational acceptances. One I have already mentioned: the unsatisfactoriness of having to admit two kinds of truth, correspondence-truth for claims about acceptances, coherence-truth for much else. It may well be thought that this constitutes a decisive objection in itself.

Not everyone will accept that, however, and for them the other problem arises from considering the remarkable agreement that there is between different people and at different times in what they believe; particularly in their observational acceptances. That we have such similar and systematically related perceptual experiences is not a matter to be accounted for by the coherence theory of truth, for it is a fact about the

²⁵ Dummett (2005: 679); cf. Dummett (2004: 53f).

²⁶ Putnam (1981: 55).

relationship of those acceptances that obtain, and truths about what acceptances obtain are correspondence-truths. The experiential beliefs of different people relate systematically to one another, and if they did not we could not share a common world or get a common language going. But *why* do they relate to one another so neatly?

If it is coherence that constitutes the truth about the physical world, the physical world cannot explain it. Of course any coherence theory of truth about the physical world can accommodate the truth of statements like “that tree is the cause of our shared visual representation of a tree”: that statement is true in virtue of the coherence of our relevant beliefs, just as other statements about the physical world are. The problem is that the explanation it gives is on the wrong level. It works all right *within* the coherent theory, but we need an explanation of the beliefs on which the theory was based, and an explanation from within the theory cannot give us that.

One might just dismiss this problem, on the grounds that nothing could in principle resolve it. Quine, whose motivation for a coherence theory of truth was again essentially semantic, insists we must stay within the coherent theory that constitutes our science, rejecting questions that try to take us beyond it: “There is no legitimate first philosophy, higher or firmer than physics, to which to appeal over physicists’ heads.” Wittgenstein just says: “It is true: there is great—and interesting—agreement here.”²⁷

Kant, I claimed, held a coherence theory of truth for the “world of appearances.” But it would recognize some truths, even about the world of appearances, that do not consist in coherence, including those about “the material conditions of experience, that is . . . sensation.” Kant recognizes that the agreement of our perceptual beliefs poses a problem that cannot be satisfactorily answered from *within* the world of appearances. To answer it he requires a reality that is wholly independent of us, a reality of “things in themselves,” which provides the ultimate ground of our sensations and so of their systematic agreement. From Kant’s own day on, many have argued that he should not have recognized things in themselves, since by doing so he seems to overstep the limits he himself laid down for knowledge, which he took to be confined to the world of appearances. They would agree with Quine and Wittgenstein, and many others too, that questions of this kind cannot properly be raised.

Yet the agreement seems to call for explanation, and an explanation that would necessarily take us outside the coherent theory. The objection to Kant arose because of the limits he placed on what can count as “knowledge.” “Knowledge” proper is possible only of the world of appearances. But there must be facts about the given element in experience, and it would seem appropriate to look for an explanation for these and for the systematic agreement amongst them, and to adopt, no doubt tentatively, the simplest and most effective. In doing this we are inevitably using our own standards. But here our standards are not constitutive of the truth, as they were within the coherent theory or within Kant’s world of appearances. They are tentative and the best we can do.

²⁷ Quine (1969: 303); Wittgenstein (1967: 50).

This is what Kant does in postulating things in themselves. He recognizes that we can have no “knowledge” about them, and he regards speculation about their nature as pointless. He does not, however, regard it as meaningless. Unlike many of his successors, and particularly the verificationists, he does not claim that our concepts are necessarily limited to the world of appearances. Or not all of them. Many are—empirical concepts like “red” and “book.” But fundamental concepts like “thing” and “cause” are not. These non-empirical concepts, the categories, can (indeed, must) be applied within our experience, and we apply them whenever we recognize things as books or see something as causing something else to happen. But they have a content that goes beyond anything that experience could ever establish, and they retain that content even when divorced from any experience actual or possible. They are sufficient to enable us to *think* about things in themselves, as entities that somehow ground our sensations. We must be able to think about such things intelligibly, for such thought is the stuff not only of the meta-physical speculations of philosophers, but of metaphysical ideas that every reflective person naturally has. It is also the kind of thinking that is indispensable (in Kant’s view) for morality and religion.

It is open to a defender of the coherence theory of truth to reject this, and to say that we cannot intelligibly formulate an answer to the question of why our observations exhibit the systematic agreement that they do. That would be to hold that the rules we use for determining the application of non-observational terms are always and necessarily built into the meanings of those terms, in such a way that they determine the conditions of their correct application and thus of the truths that contain them. The rules correspond to no reality that is independent of the system. This, as we saw, is something the verificationist has to hold, since the rules themselves are not susceptible of any kind of empirical verification; others hold it too, for similar reasons.

That would leave the agreement in our observations without any explanation except at the physical level. But someone who takes this view will say that it is only at this level that the request for explanation really makes sense.

To others it would seem more attractive to adopt a position more like Kant’s, and to hold that we can at least frame a tentative hypothesis about the source of the agreement. This requires some modification of the semantic considerations, but that might be thought to be required anyway by the thought that truths about acceptances cannot consist in coherence. For the standards and principles that we use in justifying our judgments about acceptances cannot be what determines the truth about them. This may encourage returning to the question whether there is really a very strong case for thinking that there are any contexts in which our standards *are* constitutive of the truth—except, perhaps, for specific areas in which we may already be inclined to doubt whether the nature of truth is the same as it is for descriptions of the world around us. The need to admit two kinds of truth certainly weakens the case for holding a coherence theory of truth for the physical world. It might be better to avoid having to think of two kinds of truth—coherence-truth, that holds perhaps within our theory of the external world, and correspondence-truth, that holds for statements about acceptances and perhaps for an account of why our observational acceptances fit systematically together.

It would seem simpler just to accept that truth is a matter of correspondence, and that there is therefore in our judgments always the ineradicable possibility of error—though not, of course, the sort of possibility that need worry us.

The coherence theory of truth is not dead. It may have a promising future as an account of the nature of truth in specific areas, like mathematics or ethics. I have argued that it will not do as a global theory of the nature of truth in general, but some will still be inclined to put forward a coherence theory of truth about the physical world, rather as Kant did. There remain problems with that. There is the problem of explaining the systematic agreement of our acceptances, if indeed that can be tackled intelligibly; there is the problem of the evidential base, which seems to force the coherence theorist to allow for indeterminacies in ways that run sharply against common sense. Above all, if the argument I have given is correct, coherence theorists must recognize that their theory cannot cover all truths, and that should be a serious worry for them. It means that they cannot give a unitary account of truth. And to the extent that the motivation for the coherence theory of truth lies in the thought that no sense can be made of the idea of correspondence with independent fact, this is a serious disappointment, for there must be independent facts, at least about acceptances. To that extent, its motivation has gone.

REFERENCES

- Ayer, A. J. (ed.) (1959). *Logical Positivism*. Glencoe, IL: Free Press.
- Blanshard, B. (1939). *The Nature of Truth*. London: George Allen & Unwin.
- Davidson, D. (1984). "On the Very Idea of a Conceptual Scheme," in D. Davidson, *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 183–98.
- Davidson, D. (1986) "A Coherence Theory of Truth and Knowledge," in E. LePore (ed.), *Truth and Interpretation*. Oxford: Blackwell, 307–19.
- Descartes, R. (1964–76). Meditation III, in *Oeuvres de Descartes*, ed. C. Adam and P. Tannery, revised edn. Paris: Vrin/CNRS.
- Dummett, M. A. E. (1978). *Truth and Other Enigmas*. London: Duckworth.
- Dummett, M. A. E. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Dummett, M. A. E. (2004). *Truth and the Past*. New York: Columbia University Press.
- Dummett, M. A. E. (2005). "The Justificationist's Response to a Realist," *Mind* 114: 671–88.
- Dummett, M. A. E. (2006). *Thought and Reality*. Oxford: Clarendon Press.
- Joachim, H. H. (1906). *The Nature of Truth*. Oxford: Clarendon Press.
- Kant, I. (1929). *Critique of Pure Reason*, 1st/2nd edn., trans. N. Kemp Smith. London: Macmillan.
- Neurath, O. (1932–33). "Protokollsätze," *Erkenntnis* III, trans. as "Protocol Sentences" in Ayer (ed.) (1959), 199–208.
- Olsson, E. J. (2002). "What is the Problem of Coherence and Truth?," *Journal of Philosophy* 99: 246–72.
- Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Putnam, H. (1983). *Realism and Reason*. Cambridge: Cambridge University Press.

- Putnam, H. (2001). "When Evidence-Transcendence is not Malign," *Journal of Philosophy* 98: 594–600.
- Quine, W. V. (1969). "Replies," in D. Davidson and J. Hintikka (eds.), *Words and Objections*. Dordrecht: Reidel, 292–351.
- Russell, B. (1906–7). "On the Nature of Truth," *Proceedings of the Aristotelian Society* 7: 32ff = 56ff of B. Russell (1910), *Philosophical Essays*. London: Longmans Green.
- Schlick, M. (1934). "Über das Fundament der Erkenntnis," *Erkenntnis* IV, trans. as "The Foundation of Knowledge" in Ayer (ed.) (1959), 209–27.
- Spinoza, B. (1985). *Ethics*, in E. Curley (ed.), *The Collected Works of Spinoza*, vol. 1. Princeton: Princeton University Press, 401–617.
- Strawson, P. F. (1966). *The Bounds of Sense*. London: Methuen.
- Wittgenstein, L. (1967). *Remarks on the Foundations of Mathematics*, trans. G. E. M. Anscombe. Oxford: Blackwell.
- Walker, R. C. S. (1989). *The Coherence Theory of Truth*. London: Routledge.
- Walker, R. C. S. (1998). "Bradley's Theory of Truth," in G. Stock (ed.), *Appearance versus Reality: New Essays on the Philosophy of F.H. Bradley*. Oxford: Clarendon Press, 93–109.
- Wright, C. (2000). "Truth as Sort of Epistemic," *Journal of Philosophy* 97: 335–64.
- Young, J. O. (2001). "A Defence of the Coherence Theory of Truth," *Journal of Philosophical Research* 26: 89–101.

CHAPTER 9

THE CORRESPONDENCE THEORY OF TRUTH

MARIAN DAVID

9.1 FACT-BASED AND OBJECT-BASED THEORIES

WHAT is truth? Addressing this question early in the twentieth century, Russell and Moore produced what is now regarded as the classical formulation of the correspondence theory of truth:

Thus a belief is true when there is a corresponding fact, and is false when there is no corresponding fact. (Russell 1971 [1912]: 129)

To say that this belief is true is to say that there is in the universe *a* fact to which it corresponds; and to say that it is false is to say that there is *not* in the universe any fact to which it corresponds. (Moore 1953 [1910–11]: 277)

This tells us that truth, or more precisely, the property *being true*, is a general relational property, involving a characteristic relation to some portion of reality. The relation is said to be *correspondence*; the portion of reality is said to be *a fact*. It is the second part that distinguishes this view from most of its predecessors: *fact-based* accounts of truth are rare prior to the twentieth century. At one point in the *Treatise*, Hume quickly mentions a two-pronged definition of truth, one prong of which refers to matters of fact: “Truth or falsehood consists in an agreement or disagreement either to the *real* relations of ideas, or to *real* existence and matter of fact” (*Treatise*: 3.1.1); and Mill, in a chapter on propositions, asks, “What is that which is expressed by the form of discourse called a Proposition, and the conformity of which to fact constitutes the truth of the Proposition?” (*Logic*: 1.5.1). These passages are suggestive but incidental; neither Hume nor Mill spent any time discussing the nature of truth. Ironically, the only indisputable

precursors of the classical formulation by Russell and Moore appear to be their idealist opponents at Oxford: F. H. Bradley (1922 [1883]: chs. 1 and 2) and H. H. Joachim (1939 [1906]: chs. 1 and 2). Of course, the idealists produced fact-based correspondence formulations not to defend them but to reject them. Russell and Moore, it seems, had been reverse-inspired by their idealist opponents to promote the very account of truth their opponents had formulated only to refute.

After Russell and Moore, the correspondence theory is often identified with a fact-based approach. But the theory has a lengthy history, and for most of that history *object-based* versions were the norm (cf. Künné 2003: 3.1).¹ The classical formulation prior to the twentieth century comes from Thomas Aquinas: *Veritas est adaequatio rei et intellectus* ("Truth is the equation of thing and intellect"); applying this to judgments, he says, "When a judgment is equated to that which is external in the thing, it is called true"—Aquinas also uses "*conformitas*" and "*correspondentia*," giving the latter a more generic sense; the word *res* is rather versatile, but it is clear from the texts that Aquinas is thinking of things not of facts (see *De veritate*: Q.1.A.1–3). He holds that "truth resides in the intellect," and that the intellect says what is true "by composing and dividing: for in every proposition it either applies to, or removes from, the thing signified by the subject, some form signified by the predicate" (*Summa Theologiae*: Q.16.A2; cf. Q.85.A5). This recalls Aristotle's extremely influential claim that thoughts are *likenesses* (*homoiomata*) of things (*De int.*: 16^a3), and his view that the thinking part of the soul must be "capable of receiving the form of an object; i.e. must be potentially identical in character with its object without being the object" (*De an.*: 429^a15). Somewhat surprisingly, Aristotle never actually defined truth in terms of the conceptual resources of his philosophy of mind, but it is clear that such a definition would fit right in here—a point not lost on medieval authors: Aristotle's philosophy of mind was crucial to the medieval history of the correspondence theory, especially to Aquinas (correspondence as conformity).

Authors of the modern period generally convey the impression that something very much like Aquinas's formulation is far too obvious to merit much, or any, discussion—which is not to say that they would accept his further views about what the correspondence relation amounts to.

Correspondence definitions from the modern as well as the medieval period tend to presuppose that the truth-bearing items (often taken to be judgments) have subject-predicate structure: for example,

A judgment is true if and only if it/its predicate corresponds with its object, i.e. with the object signified by the subject-term of the judgment.

Note that object-based definitions of this sort actually involve *two* relations to an object: (i) a relation holding between the subject-term of the judgment and the object the

¹ There really is no such thing as *the* correspondence theory of truth: there are numerous more or less related theories, mini-theories, and theory sketches—the definite article is used merely for expository convenience.

judgment is *about*; and (ii) a correspondence relation holding between that object and the predicate of the judgment. This second relation might then be understood in terms of the predicate of the judgment signifying a property (form, mode) that is *instantiated by* (inheres in, modifies) the object of the judgment.

Thus presupposing subject-predicate structure came naturally to modern and medieval authors, owing to the dominance of Aristotelian term logic, which was based on the assumption that the items that appear as premises and conclusions of arguments are compounds of subjects and predicates. But the presupposition makes accounts of this sort vulnerable to the objection that they are not sufficiently general because they fail to apply to items that lack the relevant structure. On the face of it, a fact-based definition seems to have the advantage here, because its formulation does not incorporate any assumptions about the structure of truth-bearing items.

9.2 TRUTH-BEARERS

Correspondence theories of truth have been given for beliefs, thoughts, ideas, judgments, statements, assertions, utterances, sentences, and propositions. It has become customary to talk of *truth-bearers* whenever one wants to stay neutral between these choices. (The term is somewhat misleading; it is intended to refer to bearers of truth *or* falsehood; or alternatively, to things of which it makes sense to ask whether they are true or false, thus allowing for the possibility that some might be neither.)

There has been much disagreement about truth-bearers. Nowadays, the main candidates are propositions, public language sentences, and sentences of the language of thought (sentential mental representations). Popular earlier candidates, *viz.* beliefs, judgments, statements, and assertions have fallen out of favor. There are two reasons for this. First, the truth-values of *logically complex* beliefs often depend on the truth-values of their simpler constituents. But there are many cases, such as conditionals (“If bats are mammals, then some mammals can fly”) and disjunctions (“The baby will be a boy or it will be a girl”), where one can hold the complex belief without believing all or any of the constituents; hence the constituents must be items that can be true or false without being believed or asserted or judged.² Second, the noun “belief” can refer to the *state* of *believing* or to *what is believed*—if the former can be said to be true or false at all (which is questionable), then only insofar as the latter, *what is believed*, is true or false.

These considerations speak against taking beliefs seriously as candidates for primary truth-bearers; and analogous considerations speak against judgments, statements, and assertions—against any candidates that contain an “assertoric” element or come with an act/object-duality.

² The point seems fundamental for a proper understanding of logic; it was made by most of the early advocates of propositions: Bolzano (1972 [1837]: I.sections 22, 34), Frege (1879: sections 2–5), Husserl (1913 [1901]: I.section 11), and Meinong (1902: section 6).

Some maintain that what is, or can be, believed is a *proposition*; they hold that propositions are the objects of the belief relation, hence the contents of belief states (and similarly for what is stated, judged, said, thought, and assumed). Others maintain that believing involves a relation to *sentences* instead, to public sentences, or more traditionally, to the inner sentences of the language of thought.

Some advocates of propositions maintain that truth and falsehood apply exclusively to propositions: saying of a sentence that it is true would then be a confusion, or an instance of metonymy, a bit like saying “The ham sandwich left without paying.” Some friends of sentences return the compliment, maintaining that propositions are never true or false because there are no propositions. Those with more ecumenical views have to say something about how “true,” when applied to bearers of one kind, relates to “true” when applied to bearers of another kind. Since it is highly implausible that these uses are entirely unrelated, one usually distinguishes between *secondary* and *primary* or *basic* truth-bearers. Secondary truth-bearers are those whose truth-values (truth or falsehood) are derived from the truth-values of primary truth-bearers, whose truth-values are not derived from any other truth-bearers. So the term “true” is ambiguous when applied to truth-bearers of different kinds, but this is not a brute ambiguity: the properties it expresses are related, with the secondary ones being definable from the primary one together with some additional relations. For example, an advocate of propositions as basic truth-bearers will hold that sentences are true or false in a secondary sense, insofar as they *express* propositions that are true or false in the primary sense. Those who still take beliefs, judgments, or assertions seriously as truth-bearers will have to draw some more connections of this sort.

When discussing theories of truth it can be crucially important to keep track of the kinds of truth-bearers to which they are supposed to apply. The standard segregation of theories into competing camps (found in textbooks, dictionaries, and handbooks) often proceeds under the assumption—really a pretense—that they are intended for the same sorts of truth-bearer taken as primary. Note that different theories designed for bearers of different kinds do not automatically compete, even when they are treated as if they did; and what looks like one theory may turn out to be two, when applied to truth-bearers of different kinds.

9.3 FOUR FORMS OF FACT-BASED THEORY

An account of truth will have a central principle, usually put forward as a *definition* of truth, in some sense of this term.³ Looking at fact-based correspondence theories, one

³ Usually as a “real definition,” which does not commit one to the claim that the core principle provides a definition in the sense of giving a synonym. Nowadays, most correspondence theorists consider it implausibly and unnecessarily bold to maintain that “true” *means the same as* “corresponds with a fact.” Interestingly, the *Oxford English Dictionary* would nevertheless seem to lend some support

finds that their central principles can take a number of different forms. The neo-classical, Russell-Moore formulation takes the following form (“iff” is short for “if and only if”):

CF: If x is a ϕ , then
 x is true iff x corresponds with some fact;
 x is false iff x does not correspond with any fact.

The letter “ ϕ ” substitutes for a noun identifying a category of truth-bearers—it reminds us that different principles result from applying the pattern to truth-bearers of different types. A specific principle exhibiting this pattern is meant to convey a universal generalization having the strength of a necessary truth; indeed, a principle of this form is typically intended as underwriting a property-identity claim, telling us that the property *being true* is the general relational property *corresponding with a fact*. (In variations on this form, the notion of correspondence may be replaced by some related notion, e.g. agreement or conformity.)

Here is a second pattern—Husserl, in his *Logical Investigations* (1913 [1901]: V) may have been an early advocate of a theory whose central principle takes this form:

CS: If x is a ϕ , then
 x is true iff x corresponds with some state of affairs that obtains;
 x is false iff x corresponds with some state of affairs that does not obtain.

Both forms invoke portions of reality that are denoted by *that*-clauses or by sentential gerundives, viz.: the fact or state of affairs *that snow is white*; the fact or state of affairs of *snow's being white*. The main difference between the two is this. CS, through its definition of falsehood, is committed to the existence of states of affairs that fail to obtain. CF has no similar commitment: there are no facts that fail to obtain, i.e. to say that a fact does not obtain means, at best, that there is no such fact, that no such fact exists.⁴ Note that some authors use “state of affairs” differently, namely as an alternative to “fact” (e.g. Armstrong 1997); they might then advocate an account of truth that employs the term “state of affairs,” like CS, but is really a notational variant of CF. This terminological issue should be distinguished from the question (which, one hopes, is a substantive one) of whether or not an account is committed to such claims as: *snow's being green* exists even though it fails to obtain.⁵

even to this bold claim, for it contains the entry: “Truth n. Conformity with fact; agreement with reality; accuracy, correctness, verity (of statement or thought).”

⁴ Strictly speaking, CF and CS, being biconditionals, are not ontologically committed to anything. Their respective commitments to facts and states of affairs arise only when they are combined with some claim to the effect that there is something that is true. The discussion assumes some such claim as given.

⁵ The difference between CF and CS is akin to the difference between property Aristotelianism, which rejects uninstantiated properties, and property Platonism, which embraces uninstantiated properties. But the relationships are not as straightforward as one might expect initially. Specifically, Aristotelianism about properties is compatible with non-obtaining states of affairs. Assume Ludwig is *not* F, but some other things are F. A property Aristotelian who subscribes to CS will then hold that the state of affairs that

The two forms are not incompatible. So advocates of CS can and will hold that their account of truth is, in effect, an analysis of the neo-classical account: facts just *are* states of affairs that obtain, they will say; moreover, for something *x* to *not* correspond with any fact just *is* for it to correspond with some state of affairs that does not obtain. On this view, CF can be reduced to CS. (Since the two forms are not incompatible, we should think of “the advocates of CF” as the ones who reject this reduction.)

The main point its advocates cite in favor of the neo-classical CF is that it is not committed to such items as non-obtaining states of affairs. They are, it is felt, not needed to account for truth and are anyway too abstract and Platonic to sustain the intuition that truth is correspondence with *reality*: hence, CF ought to be preferred over CS. One would expect this point to meet widespread approval. We will find later that CS is very popular nonetheless, though it is not usually expressed in the way I have expressed it here.

A third form that can be found in the literature is of interest primarily as a possible source of confusion. Both CF and CS should be distinguished from

C†: If *x* is a ϕ , then
 x is true iff *x* corresponds with some fact that exists;
 x is false iff *x* corresponds with some fact that does not exist.

Advocacy of a principle of this sort, if not the product of confusion, signals commitment to Meinongianism, the thesis that there are things/facts that do not exist. More frequently, it is a product of confusion, a confused version of CF or CS. Moore, for example, goes through a number of formulations very much like C† before he finally settles on CF: there are indications that he is not fully aware of the difference between the two (1953 [1910–11]: 255–6, 267, 269, 277). At times, it can be quite difficult to detect which of these forms a given author actually wants to embrace. Wittgenstein will serve as a prominent illustration of this. In the *Tractatus* (1961 [1921]) we find: “If an elementary proposition is true, the [corresponding] state of affairs exists: if an elementary proposition is false, the [corresponding] state of affairs does not exist” (4.25). This looks like Wittgenstein is proposing a notational variant of C†, with “state of affairs” replacing “fact.” But when we check the German original, we find him saying that, if the elementary proposition is true, the state of affairs obtains (*besteht*), if it is false, the state of affairs does not obtain (*besteht nicht*). This looks very much like an instance of CS.⁶ But then again, there are other passages that imply that “state of affairs” is just Wittgenstein’s term for (atomic)

Ludwig *is* F exists but does not obtain. Divergence from property Platonism comes out only if neither Ludwig nor anything else is F: in this case, the Aristotelian will say that there is no such state of affairs as Ludwig’s being F (because there is no such property as being F), while the Platonist will still hold that there is such a state of affairs, albeit one that fails to obtain.

⁶ Confusingly, he says that propositions (*Sätze*) represent *the obtaining* and *non-obtaining* of states of affairs (cf. 4.1), and tends to say that a fact is *the obtaining* of a state of affairs (e.g. 2.062), even though he is clearly committed to the view that a fact is a state of affairs that obtains (cf. 2.04 and 1.1).

facts; indeed, he approved of an earlier translation (1922) that renders *Sachverhalt* as “atomic fact”; and in response to a query by Russell, he said: “*Sachverhalt* is what corresponds to an Elementarsatz if it is true” (1995 [1919]: 125), which suggests a version of CF rather than CS.

A fourth simple pattern was popular for a time (cf. Russell 1956 [1918]: I and III; Broad 1933: IV.2.23; Austin 1950: 129) but seems to have fallen out of favor. It is better expressed using the notion of agreement in place of correspondence:

$C \rightleftharpoons$ If x is a ϕ , then
 x is true iff x agrees with some fact;
 x is false iff x disagrees with some fact.

This formulation promises an account of falsehood that avoids CS’s commitment to non-obtaining states of affairs and C^\dagger ’s commitment to non-existent facts by invoking a relation of mis-correspondence. The proposal gives rise to some obvious questions: Which fact or facts does a given falsehood disagree with? and, What keeps a truth, which by definition agrees with some fact, from also disagreeing with some other facts? The idea must be that each truth and falsehood agrees or disagrees with exactly one fact: if it is a fact that Ludwig is hungry, then “Ludwig is hungry” is true because it agrees with that fact, and “Ludwig is not hungry” is false because it disagrees with that same fact; if, on the other hand, it is a fact that Ludwig is not hungry, then “Ludwig is hungry” is false because it disagrees with, and “Ludwig is not hungry” is true because it agrees with, the fact that Ludwig is not hungry. It is natural to understand the two relations that are involved here as species of one genus, as Russell apparently did in 1918: “The essence of a proposition is that it can correspond in two ways with a fact, in what one may call the true way or the false way” (1956 [1918]: 208). An advocate of the more traditional CF is likely to point out that $C \rightleftharpoons$ is at a disadvantage compared to CF because it introduces a troublesome second relation that does not appear to be needed—the absence of the relation responsible for truth will serve to account for falsehood—while no advantage of $C \rightleftharpoons$ is discernible.⁷

We have looked at four forms of core principle or definition for fact-based correspondence theories. In the following, I will set the latter two aside and treat CF and CS as paradigmatic. Moreover, I will at times (but not always) gloss over the difference between these two and condense the correspondence approach into the formula “truth is correspondence with a fact”; advocates of CS can agree with this, since they hold that obtaining states of affairs *are* facts.

⁷ $C \rightleftharpoons$ may be a child of the idea that facts are the meanings (denotations) of sentences; see Wittgenstein (1961a [1913]: 94): “Thus a proposition has two poles (corresponding to case of its truth and case of its falsity [sic]). We call this the *sense* of a proposition. The *meaning* of a proposition is the fact which actually corresponds to it. The chief characteristic of my theory is: *p* has the same meaning as *not-p*.”

9.4 NON-IDENTITY, CORRELATION, AND STRUCTURE

Some correspondence theories of truth are one-liner mini-theories, consisting of little more than a specific version of CF or CS. Normally, one would expect a bit more, even from a philosophical theory.⁸ One would expect a correspondence theory to discharge a triple task: it should tell us about the workings of the correspondence relation, about the nature of facts, and about the conditions that determine which truth-bearers correspond with which facts.

We can approach this by considering some claims a correspondence theorist might want to add to her central principle to flesh out her theory. The first such claim says that the correspondence relation must not collapse into identity—"It takes two to make a truth" (Austin 1950: 118):

Non-identity: No truth is identical with a fact correspondence with which is sufficient for its being a truth.⁹

A correspondence theorist will take this for granted as constitutive of the very idea of a correspondence theory—those who advance contrary arguments to the effect that correspondence does collapse into identity regard such arguments as refutations of any form of correspondence theory (cf. Moore 1993 [1901–02]; Frege 1918–19: 60).

Concerning correspondence, two separable aspects of this notion can be distinguished: *correspondence as correlation* and *correspondence as congruity* or *isomorphism* (cf. Pitcher 1964: 9–14). Pertaining to the first aspect, familiar from mathematical contexts, a correspondence theorist is likely to adopt claim (a), and some may even adopt both, claims (a) and (b), of:

Correlation: (a) Every truth corresponds with exactly one fact.
 (b) Different truths correspond with different facts.

Taken together, (a) and (b) say that correspondence is a one–one relation. This seems to be a needlessly strong claim, and it is not easy to find a real-life correspondence theorist who explicitly embraces part (b): Why shouldn't different truths correspond

⁸ But consider that mini-theories are by no means unheard of in philosophy. Incidentally, the frequent occurrence of the term "theory" in discussions of truth should not be taken too seriously: it owes its prevalence primarily to the dearth of other handy terms—no one thinks that a "theory" of truth has the status of a genuine scientific theory.

⁹ It would be less cumbersome to say simply that no truth-bearer is identical with a fact. However, Wittgenstein, for one, thought that a proposition (*Satz*) is a fact, though not the sort of fact that would be sufficient for the truth of the proposition (1961 [1921]: 3.14).

with the same fact, as long as they are not too different?¹⁰ Explicit commitment to (a) is also quite rare. However, correspondence theorists tend to move comfortably from talk about a given truth to talk about *the* fact it corresponds with—a move that signals commitment to (a).

Correspondence as correlation does not imply anything about the inner nature of the corresponding items. Contrast this with *correspondence as congruity* or *isomorphism*: it requires the corresponding items to have the same or sufficiently similar *structure*. This aspect of correspondence, which is more prominent (and more notorious) than the previous one, is also much more difficult to make precise. Let us say, rather roughly, that a correspondence theorist may want to add a claim to her theory committing her to something like the following—letting “ ϕ ” again take the place of a noun specifying a type of truth-bearer:

Structure:

If a ϕ (a truth-bearer) corresponds with a certain fact, then they have the same or sufficiently similar structure: the overall correspondence between a true ϕ and a fact is a matter of part-wise correspondences, of their having corresponding constituents in corresponding places in the same structure, or in sufficiently similar structures.

Assume, if only for the sake of illustration, that sentences are (basic) truth-bearers. A correspondence theorist is apt to say that “Ludwig is hungry,” if true, corresponds with the fact that Ludwig is hungry; and one who is also committed to Structure will add that the sentence and the fact have constituents arranged in a certain way. The constituents of the sentence are the name “Ludwig” and the predicate “is hungry” or “hungry.” The constituents of the fact are Ludwig and the property *being hungry*. The name corresponds with Ludwig, and the predicate corresponds with the property *being hungry*. Proposition and fact have the same structure and have corresponding constituents in the same structure. Of course, this illustration of the workings of Structure applies only to the most simple sentences; more complicated accounts have to be devised for sentences with more complicated structures.

Correlation and Structure reflect distinct aspects of correspondence. One might want to endorse the former without the latter; though it is hard to see how one could endorse the latter without embracing at least part (a) of the former.¹¹

A common objection to correspondence theories targets specifically the correspondence relation, charging that it is unintelligible, obscurantist, mysterious, and so on. Structure promises a response to objections of this sort; the details of the response will depend on the choice of primary truth-bearers.

¹⁰ Moore is the only one I am aware of who explicitly advocates both (a) and (b), though at later places in the same work he mentions only (a); cf. Moore (1953 [1910–11]: 256, 268, 276–7).

¹¹ Austin (1950) advocates correspondence with Correlation but without Structure; he does not tell us whether he wants both parts of Correlation or only part (a).

If *sentences* are taken as primary truth-bearers (tokens of declarative sentence types), the correspondence relation will reduce to underlying semantic relations between words and objects (and properties and relations); relations such as reference, denotation, satisfaction, representation, expression, etc.¹² Consequently, a correspondence theory of this sort will be a spin-off from semantics. Essentially the same point holds if the primary truth-bearers are taken to be the *mental sentences* of the language of thought (sentential mental representations): a correspondence theory of this sort will be a spin-off from psycho-semantics—the theory of intentionality construed as incorporating a representational theory of the mind (cf. Fodor 1989). This puts “mystery”-objections to correspondence into perspective. Given sentences as primary truth-bearers, the correspondence relation is no more—but no less—mysterious than semantic relations in general. Sure enough, such relations have some curious features, and they raise difficult questions—most notoriously: Can they be explained in terms of natural (causal) relations, or do they have to be regarded as irreducibly non-natural aspects of reality? Some philosophers, taking the answer to this question to be negative, have claimed that semantic relations are too mysterious to be taken seriously. But one should bear in mind that this is a very general and extremely radical rejection of semantics as a whole, of the very idea that words can be *about* things. The common practice to aim this attack specifically at the correspondence theory is misleading. As far as the intelligibility of the correspondence relation is concerned, a correspondence theory with sentences as primary truth-bearers will stand, or fall, with the general theory of reference and intentionality.

But what about a correspondence theory taking *propositions* as primary truth-bearers? It depends on one's view of the nature of propositions. On a broadly *Fregean* view, propositions are composed of *concepts*—where a concept is construed not as a psychological item but as an objective way of conceiving of things or properties or relations.¹³ On this view, the resulting picture is similar to the one above, since the relations between concepts and the things and properties they are concepts of are similar to semantic relations. A correspondence theory of this sort will be a part of concept-semantics—of course, advocates will have to motivate embracing Fregean concept-propositions in the first place, but that is an issue that is largely independent of the details of one's theory of truth. On the so-called *Russellian* view, propositions are constituted, not of concepts of objects and properties, but of the objects and properties themselves. If propositions, thus understood, are taken as primary truth-bearers, the picture changes considerably; indeed, it changes so much that objections aimed at the correspondence relation between propositions and facts become irrelevant (see section 9.5).

¹² But the sense of a sentence must not be identified with a corresponding fact, on pain of the absurd consequence that false sentences have no sense. Tokens are preferable to types, because of sentences with context-sensitive elements, e.g. “He is hungry now.” The type is neither true nor false; a token of this type is true or false, depending on the time and on the referent of “he” in the context in which the token is produced.

¹³ Frege himself talked of *modes of presentation* or *senses* instead; he used “concept” (*Begriff*) in a different and somewhat strange way.

Embracing Structure without qualifications would require a perfect isomorphism between true truth-bearers and facts. Critics, looking at the vagaries of our language, find this a ludicrous idea. It has never been seriously advocated, not for ordinary language. Instead, proponents try to isolate the “relevant” structural features of our sentences, aiming to uncover their *logical form*, which is then exhibited by way of an *ideal language*—often the language of predicate logic. In a setting with mental sentences or propositions, the formulas of the ideal language are supposed to represent these primary truth-bearers, offering a reliable guide—ideally a perfect guide—to their inner structure. In a setting without such intermediaries, the true formulas of the ideal language will be taken to directly represent reality, their grammatical structure mirroring the ontological structure of facts.

The following advantages may be claimed for this ideal-language strategy: (i) Since different ordinary sentences might be mapped onto the same ideal formula, it allows one to stay away from part (b) of Correlation by assigning the same fact to structurally different but synonymous ordinary sentences, for example to active/passive pairs. (ii) It allows one to avoid assigning odd objects to parts of ordinary sentences that are irrelevant to truth-value e.g. to discourse indicators such as: “accordingly,” “nevertheless,” “if I may say so,” etc. (iii) It allows one to avoid assigning anything to parts of ordinary sentences that are relevant to truth-value but would require questionable ontological correlates such as “the average husband,” “the present King of France,” “the sake of,” etc.

On the whole it seems clear that Structure does not make life easy for correspondence theorists: they want as much of it as they can get while steering clear of the claim that each and every wrinkle of our true sentences faithfully mirrors an aspect of the structure of reality.

Structure seems to have an important consequence bearing on the comparison between fact-based and object-based approaches. The former identify *being true* with what appears to be a relatively simple property: the general relational property *corresponding with some fact*. Using the language of first-order predicate logic, the form of this property can be displayed thus—say x is a true sentence:

x is such that $(\exists y)(xRy \ \& \ Fy)$

Now consider an object-based approach: “Ludwig is hungry” is true, if and only if, there is some object named by “Ludwig,” some property expressed by “hungry,” and the object instantiates the property. Note that this account applies only to this one sentence; generalizing requires something like:

If x is a subject-predicate sentence, then

x is true iff, there is an object y , a property z , a term t_1 , and a term t_2 , such that (i) t_1 is the subject-term of x , and t_2 is the predicate-term of x , (ii) t_1 names y , and t_2 expresses z , and (iii) y instantiates z .

Evidently, the general relational property that is ascribed here to a true subject-predicate sentence x is much more complex than the one ascribed by a fact-based theory—its form comes out as:

$$x \text{ is such that } (\exists y)(\exists z)(\exists t_1)(\exists t_2)(Oy \ \& \ Pz \ \& \ Tt_1 \ \& \ Tt_2 \ \& \ (t_1R_sx \ \& \ t_2R_px \ \& \ (t_1R_no \ \& \ t_2R_ez \ \& \ (yR_iz))))$$

Moreover, this works only for the simplest subject-predicate sentences; true sentences of other grammatical forms will require other (more complex) properties. This multiplicity raises some questions. Is it that there is no such thing as *the* property *being true* (not even for sentences), instead there is a large number of different properties? Or is there a single property but it is highly disjunctive? Or is there some way of providing for unification amidst this multiplicity? Either way, the account faces the worry that it is not sufficiently general, because true sentences of some grammatical form or other have been left out.¹⁴

Prima facie, a fact-based approach seems to enjoy an advantage here (as was pointed out in the first section), because its core principle does not make any explicit assumptions about the structure of the truth-bearing items. However, as soon as an advocate of this approach adds Structure to her theory, this advantage would seem to be neutralized. Her account will get unwieldy at the level where it specifies the conditions under which x corresponds with y , for variously structured truth-bearers x and variously structured facts y . The complications that arise here for the fact-based account will mirror closely the complications that arise for an object-based account, the sole difference being that the former encounters them at one remove, that is at the level of the correspondence relation, while the latter encounters them right away. It seems then that, despite first appearances to the contrary, simplicity considerations will not speak much in favor of a fact-based approach—at least not if it wants to embrace Structure.

9.5 STRUCTURE AND TRUTH-BEARERS

Propositions appear to be the most popular candidates for primary truth-bearers nowadays.¹⁵ But there are two importantly different and competing views about the nature of propositions: Fregean versus Russellian. In this section, I want to explain why

¹⁴ Some medieval semanticists, e.g. Buridan, offer lengthy lists of truth-conditional clauses for sentences of different grammatical categories, aiming to cover all of them (cf. *Sophismata*: 858f). Buridan does not discuss how many “truth properties” there are on his account.

¹⁵ Mental sentences were the preferred primary truth-bearers throughout the medieval and the modern period; they were neglected in the first half of the twentieth century but made a comeback in the second half through the revival of the representational theory of the mind. Somewhat confusingly (to us), for many centuries the term “proposition” (*propositio*) was reserved exclusively for sentences, spoken, written, and mental; this use was made official by Boethius in the sixth century and is still found in Mill’s *Logic* in 1843. Some time after that, e.g. in Moore’s (1901–02), “proposition” switched sides, the

the outcome of this debate will make a significant difference to the correspondence approach to truth.

Assume, first, that Frege was right about the nature of propositions. If so, then propositions—he called them “thoughts” (*Gedanken*)—are abstract, mind- and language-independent necessary beings, composed of abstract, mind- and language-independent concepts.¹⁶ Advocates of a correspondence theory for Fregean propositions are apt to hold that there is a fairly good match between the structure of a true proposition and the structure of a fact it corresponds with, but that the match is not (always) so tight that it precludes different true propositions from corresponding with the same fact, as long as they are not too different—a prominent example: the proposition that Cicero is an orator differs, on this view, from the proposition that Tully is an orator, the first having the *Cicero*-concept as a constituent where the second has the *Tully*-concept as a constituent; nevertheless, our correspondence theorist will hold that these propositions correspond with the same fact, the fact that Cicero, i.e. Tully, is an orator. On this view, the core principle

CF: If x is a ϕ , then
 x is true iff x corresponds with some fact,
 x is false iff x does not correspond with any fact,

is designed for propositions. Sentences, including mental sentences (if any), are secondary truth-bearers. Correspondence theorists of this stripe are likely to hold that the grammatical structure of sentences is a fallible guide to the structure of the propositions they express. So they will employ the ideal-language strategy (see above); and they will design the formulas of their ideal language to mirror faithfully the structure of the propositions they express. In this way, the ideal language is supposed to provide—via the correspondence between true propositions and facts—a fairly faithful representation of the real structure of facts.

A theory of this sort will aim at a *double-correspondence* account of truth. True propositions correspond with facts. In addition, true sentences are also said to correspond with facts, though this is a derived, composite notion of correspondence: for a sentence to correspond with a fact is for it to stand in some relation R to a proposition which corresponds with a fact in the primary way. Since R is a relation between truth-bearer and truth-bearer, it will differ from the primary correspondence relation; nevertheless, it will raise some analogous issues: whether it is one–one or many–one; whether it involves an isomorphism and, if yes, how tight the isomorphism is supposed to be.

term now being used for *what is said* by uttering a sentence, for what is believed, judged, stated, assumed, etc.—with occasional reversions, e.g. by Russell.

¹⁶ The “composed of” is very much *not* to be taken to imply that propositions have to be put together by some process or agent. I should mention that Frege himself rejected any form of correspondence theory for propositions; cf. his (1918–19: 59–60).

I have presented this approach as centered on the neo-classical correspondence pattern CF; variants can be developed around the alternative pattern CS. We have seen that Wittgenstein may have been thinking along CS lines, if only sometimes and maybe confusedly. Husserl is plausibly interpreted as advocating CS with Fregean propositions as primary truth-bearers (cf. 1913 [1901]: I and V, sections 28, 33, 36); he called propositions *Sätze* and *ideale Bedeutungen* (“ideal meanings”).

Assume now that early Russell was right about propositions. Unlike Frege, for whom all propositions are made from the same cloth (from concepts), early Russell is very liberal about the constituents of propositions. There is one requirement: every proposition must contain at least one n -place relation (properties count as 1-place relations), together with some other constituent or constituents that, as one says, plug into the slot or slots of the relation. There is no restriction on what sorts of items can combine with the n -place relation to constitute a proposition. Purely general propositions are composed of an n -place relation together with more properties and relations (or functions). A *singular proposition* is composed of an n -place relation together with the particular thing or things the proposition is about, such as a bicycle, or Arnold Schwarzenegger, or a mountain, or numbers, or dogs.¹⁷

Note, there is no room here for a correspondence theory of truth for propositions. Consider the Russellian proposition that Ludwig is hungry. It is composed of Ludwig and the property of being hungry; assume this proposition is true: it must *be* a fact, the fact that Ludwig is hungry. It is hard to see how a true proposition of this sort could not be a fact: what would a fact *be*, if not this sort of thing? So the principle of Non-identity is rejected, and with it goes the correspondence theory of truth: “a truth differs in no respect from the reality with which it was supposed merely to correspond” (Moore 1901–02: 717). What we find here instead is an *identity theory*: a proposition is true iff it is a fact, and false iff it is not a fact.¹⁸

Do Russellians renounce the correspondence theory of truth entirely? Not at all. Sure enough, they have no room for CF with propositions as primary truth-bearers. But correspondence will enter into their account of truth for sentences, public or mental. This account takes the form CS, where “ ϕ ” stands for any category of truth-bearers other than Russellian propositions:

CS: If x is a ϕ , then
 x is true iff x corresponds with some state of affairs that obtains;
 x is false iff x corresponds with some state of affairs that does not obtain.

¹⁷ Russell himself believed in Russellian propositions only for a relatively short time (1899–1907); later he used the term “proposition” to refer first to public sentences (1956 [1918]) then to mental sentences (1956a [1919]).

¹⁸ Cf. Russell (1994 [1905]). Russell and Moore turned to the correspondence theory only after they had rejected Russellian propositions. This is why the neo-classical formulations at the beginning of this chapter are formulated for *beliefs*: at the time, Russell and Moore were seriously committed to the view that belief *states* are the primary truth-bearers.

Russellian propositions show up here in the guise of states of affairs that obtain or fail to obtain. Commitment to propositions *and* states of affairs tends to be regarded with some scorn, as a gratuitous ontological duplication: more than too much by way of dubious ontology. But Russellians are not committed to states of affairs *in addition* to propositions, for propositions, on their view, must already be states of affairs. This conclusion is well-nigh inevitable, once true propositions have been identified with facts. If a true proposition is a fact, then a false proposition that might have been true would have been a fact, if it had been true. So, a (contingent) false proposition must be the same kind of being as a fact, only not a fact—an unfact; but that just is a non-obtaining state of affairs under a different name. Russellian propositions are states of affairs: the false ones are states of affairs that do not obtain, and the true ones are states of affairs that do obtain.¹⁹

The Russellian view of propositions is popular nowadays; moreover, contemporary Russellians hold that the structure of sentences (public or mental) corresponds rather closely with the structure of propositions—at least in basic cases. The theme was sounded by David Kaplan:

If I may wax metaphysical in order to fix an image, let us think of the vehicles of evaluation—the what-is-said in a given context—as propositions. Don't think of propositions as sets of possible worlds, but rather as structured entities looking something like the sentences which express them. For each occurrence of a singular term in a sentence there will be a corresponding constituent in the proposition expressed . . . In general, the constituent of the proposition will be some sort of complex, constructed from various attributes by logical composition. But in the case of a singular term which is directly referential, the constituent of the proposition will be the object itself. (Kaplan 1989: 494)

Russellians are typically committed to a correspondence theory for sentences and mental sentences (if any). Their theory typically includes a fairly bold version of Structure and its central principle takes the form CS—though states of affairs are often called “propositions.”²⁰

Curiously, contemporary Russellians hardly ever refer to propositions as facts or states of affairs.²¹ The reason, I think, is that they are much concerned with understanding belief, belief attributions, and the semantics of sentences. In such contexts, it is more natural to talk proposition-language than state-of-affairs-language. It feels odd (wrong) to say that someone believes a state of affairs, or that states of affairs are true or false. For that matter, it also feels odd (wrong) to say that some propositions are facts, that facts

¹⁹ Frege held that facts are true Fregean propositions: “‘Facts, facts, facts’ cries the scientist if he wants to bring home the necessity of a firm foundation for science. What is a fact? A fact is a thought that is true” (1918–19: 74). But a Fregean about propositions does not have to hold this view. On the contrary, she can advocate a correspondence theory instead. Indeed, most friends of facts would hold that Fregean concept-propositions are too fine-grained and not worldly enough to be plausible candidates for facts and states of affairs; for dissenters, see Chisholm (1976) and Dodd (2000).

²⁰ Cf. Russell (1903: ch. 4); Salmon (1986: ch. 2); and Soames (1987: section 6).

²¹ But see Wettstein (1986).

are true, and that propositions obtain or fail to obtain. Nevertheless, all of this must be the literal truth, according to the Russellians. They have to claim that “proposition” and “state of affairs,” much like “evening star” and “morning star,” are different expressions for the same things—though they come with different associations and are at home in somewhat different linguistic environments, which accounts for the felt oddness when one expression is transported to the other’s environment.

9.6 LOGICAL ATOMISM

Given the great variety of truth-bearers, a correspondence theory is committed to all sorts of complex facts, such as disjunctive, conditional, negative, universal, probabilistic, subjunctive, and counterfactual facts. Moreover, since it assigns corresponding entities to all (relevant) constituents of truth-bearers, it is committed to assigning “logical objects” to the logical constants (“or,” “if-then,” “not,” etc.). Many philosophers have found it hard to believe in the existence of all these funny facts and objects.

Logical atomism was designed to address the “funny fact”-objection. Remember the grammar of a language for propositional logic: if “p” and “q” are well-formed, then so are the forms “not-p,” “p & q,” “p or q,” “if p then q”; and remember the rules for evaluating complex formulas, enshrined in the truth-tables. They can be understood (a) as tracing logical relations between complex facts and simpler facts, or (b) as exhibiting how the truth-values of complex sentences can be explained in terms of their logical relations to simpler sentences and the correspondence or non-correspondence of simple “atomic” sentences with “atomic” facts. Logical atomism takes option (b); in its pure form, it aims to avoid all commitment to funny facts and funny objects:

Whatever corresponds in reality to compound propositions [*Sätze*] must not be more than what corresponds to their several atomic propositions. (Wittgenstein 1961a [1913]: 100)

You must not look about the real world for an object which you can call “or,” and say “Now look at this. This is ‘or’.” (Russell 1956 [1918]: 209f)²²

Note that atomism is *not* for the friends of Russellian propositions. The argument is straightforward. We have logically complex beliefs some of which are true. According to the friends of Russellian propositions, the contents of our beliefs are Russellian

²² Bradley is again a precursor: “What fact is asserted in negative judgments? Has every negation I choose to invent a real counterpart in the world of things? . . . Consider again hypothetical judgments. *If* something is, *then* something else follows, but should neither exist, would the statement be false? It seems just as true without facts as with them, and, if so, what fact can it possibly assert? The disjunctive judgment will again perplex us. ‘A is *b* or *c*’ must be true or false, but how in the world can a *fact* exist as that strange ambiguity ‘*b* or *c*’? We shall hardly find the flesh and blood alternative that answers to our ‘or’” (1922 [1983]: ch. 1.6).

propositions, and the contents of our true beliefs are true Russellian propositions. Since true Russellian propositions are facts, there must be at least as many complex facts as there are true beliefs with complex contents (and at least as many complex states of affairs as there are true or false beliefs with complex contents). Atomism may work for sentences, public or mental, and for Fregean propositions; but not for Russellian propositions.²³

Strictly speaking, atomism *rejects* the correspondence theory. It denies that *being true* is *corresponding with some fact* and holds that there are many truths that do *not* correspond with any facts: CF must be confined to atomic truth-bearers. However, since atomism aims to explain truths that don't correspond with facts in terms of their logical relations to atomic truths that do correspond with facts, it is usually counted as a version of the correspondence theory. Atomists have not discussed the question how many "truth properties" there are, on their view. If they maintain that there is a single property *being true*, they have to allow for this property to be very complex.

Logical atomism has rarely been held in a pure form. Wittgenstein (1961 [1921]) accepts negative facts, but wants to get by without general facts; he wants to analyze universal and existential generalizations in terms of infinite conjunctions and disjunctions of atomic truth-bearers. Russell (1956 [1918]: 235) opposes this; he ends up accepting negative facts, existential facts, and universal facts, but rejecting disjunctive and conditional facts. Armstrong (1987) rejects negative facts and tries to get by with only one universal fact, albeit an extremely general one. In practice, it can be difficult to tell whether a view is a more or less pure form of atomism, because there are two competing options for defining the notion of an atomic sentence: (i) as a singular sentence composed of n names and an n -place predicate; (ii) as a sentence that has no sentence as a proper constituent. Generalizations come out as atomic on (ii) but as non-atomic on (i). The status of negations of singular sentences is doubtful either way; it depends on whether one takes the scope of negation to be wide ("Not: Ludwig is hungry") or narrow ("Ludwig is-not hungry").

In addition to differences about fundamentals, atomism must face the objection that it cannot avoid commitment to funny facts in any case. For there are molecular truth-bearers, such as subjunctive and counterfactual conditionals, that tend to provoke the funny-fact objection but appear to be immune to the atomists' treatment, because their truth-values are not determined by the truth-values of their constituents. Belief attributions are also a cause for concern: "Tina believes that Ludwig is hungry" appears to be molecular; but it is far from clear how its truth-value could be explained in terms of atomic constituents.

The atomist's strategy of accounting for the complex in terms of the simple can be pushed further by going *subatomic*. Such an account analyzes truth-bearers, say sentences, into their subsentential constituents and the correspondence relation into

²³ Russell advocated atomism in 1918, some time after he had rejected Russellian propositions—on the grounds that a person with "a vivid sense of reality" couldn't believe that there are such things as objective falsehoods; see his (1956 [1918]: 223).

appropriate semantic subrelations: names *refer* to objects; predicates (open sentences) apply to, or are *satisfied* by objects. Satisfaction of complex predicates is handled recursively in terms of logical structure and satisfaction of simpler constituent predicates. The recursions are anchored in a *base-clause* addressing the satisfaction of *primitive* predicates. Truth for singular sentences, consisting of a name and an arbitrarily complex predicate, is defined thus: “A singular sentence is true iff the object referred to by the name satisfies the predicate.” Logical machinery provided by Tarski (1935) can be used to turn this sketch into a more general account—one that handles sentences containing relational predicates and quantifiers, and covers truth-functional molecular sentences as well.

Subatomism constitutes a return to an object-based approach to truth. Opponents of facts favor it, for it promises to get by without assigning any sentence-like slices of reality to truth-bearers, not even to atomic truth-bearers. There has been considerable debate over whether subatomism should still count as a version of the correspondence theory at all. The issue would seem to depend on how the reference relation and the base-clause for satisfaction are handled: Tarski gave the latter a “deflationary” form, roughly: an object y satisfies a primitive predicate t iff (t = “is hungry” and y is hungry) or (t = “is a dog” and y is a dog) or . . . and so on, until the store of predicates of the language under study has been exhausted. Others would argue for a more “substantive” base-clause, taking us back to a traditional object-based correspondence account: an object satisfies a primitive predicate iff the predicate expresses some property that is instantiated by the object.²⁴

9.7 FACTS

Some philosophers are much opposed to taking facts seriously; others find it obvious that there are facts. To some extent this debate may be rooted in a disagreement about what we *perceive*. Advocates of facts take it to be evident that many facts are observable: when one looks at a cat that is on a mat, then one doesn’t just see a cat on a mat, one sees *that* the cat is on the mat (and if there is a dog right next to it but not on the mat, some would even say one can see *that* the dog is *not* on the mat). Foes of facts take it to be evident that we do not *see* such *that*-ish items (and certainly not negative ones).

Davidson (1969) argues that the correspondence theory—which he identifies with a fact-based theory—is bankrupt because it cannot avoid the trivializing consequence that all true sentences correspond with the same fact: the Big Fact. His argument, the so-called *slingshot*, has been criticized repeatedly. Without going into details, I simply

²⁴ Compare e.g. the discussion in Field (1972). It turns out that relational predicates make it necessary to talk of predicates being satisfied by *ordered sequences* of objects rather than by the ordinary objects themselves (cf. Tarski 1935: section 3). Davidson (1969) maintains that satisfaction by ordered sequences is all that remains of the traditional idea of correspondence with facts.

report the result of a book-length study by Stephen Neale (2001). Fact theorists who subscribe to Russell's analysis of definite descriptions are immune to the slingshot, assuming they hold, as they do, that, for example, the fact that Cicero is an orator \neq the fact that the denouncer of Catilina is an orator, on the grounds that the former is a singular fact while the latter is a general fact, the two sharing only the constituent *being an orator*. Those who do not subscribe to Russell's analysis will still sidestep the slingshot, provided they *reject* the following two claims: all logically equivalent truths correspond with the same fact; and, the fact that a is F = the fact that the x , such that x is identical with a , is F .

Any fact-based correspondence theory committed to Structure is thereby committed to the idea that facts are structured complexes composed of parts or constituents. Some might question whether this kind of language can really be taken seriously. The worry is that part-whole talk has its natural home in the domain of material things and is overstretched when applied to the domain of facts. Here are two principles about "the constitution of facts" that might be cited to substantiate this worry, because they seem rather peculiar.

First, different facts can have the same constituents at the same time, *viz.* the fact that Ludwig bites Tina, and the fact that Tina bites Ludwig. Sure enough, different material things can also have the same parts, but not, it seems, at the same time—and if you point out that a statue and the lump of clay it is made from may raise a problem for the latter principle about material composition, you will acknowledge that things are still rather different when it comes to facts: the two facts mentioned above are like *two statues* made from the same clay existing at the same time.

Second, it seems that the same constituent can "occur" in the same fact more than once; for example the fact that Ludwig is Ludwig, that Ludwig bites Ludwig, that Ludwig is a son of the father of Ludwig, and so on.

These principles do indeed seem a bit odd. But note that they apply not just to facts; they apply to states of affairs in general, that is, to Russellian propositions, and they apply to Fregean propositions. If they indicate that there is something wrong with thinking of facts as having constituent structure, then they indicate that there is also something wrong with thinking of propositions as having constituent structure. In this connection one might find it interesting to consider that these principles *also* apply to the constitution of sentence types.

There are two more questions in this vicinity. The first is known as "the problem of the unity of the proposition": What distinguishes the Russellian proposition, i.e. the state of affairs that Ludwig bites Tina, from the set of its constituents: {Ludwig, Tina, biting}? One might be tempted, at first, to respond that in the state of affairs the biting relation relates Ludwig to Tina. But a response of this sort would be a mistake. It would imply that there are no states of affairs that fail to obtain, that is, no Russellian propositions that are false.

Finally, the fact (or the state of affairs) that Ludwig bites Tina differs from the fact that Tina bites Ludwig. But how do they differ? This is not obvious, for it seems that they have the same constituents: Ludwig, Tina, biting, and the same form: xRy . To say that in the first Ludwig occurs to the left and Tina to the right of the biting relation while in the

second Ludwig occurs to the right and Tina to the left, confuses the facts with the sentence tokens that are spread out on this piece of paper.

One may note that both questions can also be raised, in slightly modified form, for Fregean propositions—and also for sentence types.

Russell worried about these questions repeatedly (e.g. 1903; 1984 [1913]); he never found an answer that satisfied him.

REFERENCES

- Aristotle. (2001). *De interpretatione, De anima*, ed. R. McKeon, *The Basic Works of Aristotle*. New York: The Modern Library.
- Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Austin, J. L. (1950). "Truth," reprinted in *Philosophical Papers*, 3rd edn. Oxford: Oxford University Press, 117–33.
- Bolzano, B. (1972) [1837]. *Wissenschaftslehre*. Seidel: Sulzbach. Partially translated as *Theory of Science*, ed. Rolf George. Berkeley: University of California Press.
- Bradley, F. H. (1922) [1883]. *The Principles of Logic* Oxford: Oxford University Press.
- Broad, C. D. (1933). *Examination of McTaggart's Philosophy*, vol. 1. Cambridge: Cambridge University Press.
- Buridan, J. (2001). *Sophismata*, in *Summulae de dialectica*, trans. G. Klima. New Haven: Yale University Press.
- Chisholm, R. M. (1976). *Person and Object*. London: George Allen & Unwin.
- Davidson, D. (1969). "True to the Facts," *Journal of Philosophy* 66: 748–64.
- Dodd, J. (2000). *An Identity Theory of Truth*. New York: St. Martin's Press.
- Field, H. (1972). "Tarski's Theory of Truth," *Journal of Philosophy* 69: 347–75.
- Fodor, J. (1989). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: The MIT Press.
- Frege, G. (1879). *Begriffsschrift*. Halle: Louis Nebert.
- Frege, G. (1918–19). "Der Gedanke: Eine logische Untersuchung," in *Beiträge zur Philosophie des deutschen Idealismus* 1, 58–77. English translation: "Thoughts," in *Collected Papers* (1984), ed. B. F. McGuinness. Oxford: Blackwell. (References are to original pagination.)
- Hume, D. (1978) [1739–40]. *A Treatise of Human Nature*. Oxford: Clarendon Press.
- Husserl, E. (1913) [1901]. *Logische Untersuchungen*. Tübingen: Niemeyer.
- Joachim, H. H. (1939) [1906]. *The Nature of Truth*. Oxford: Oxford University Press.
- Kaplan, D. (1989). "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press, 481–563.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Clarendon Press.
- Meinong, A. (1902). *Über Annahmen*. Leipzig: J. A. Barth.
- Mill, J. St. (1996) [1843]. *A System of Logic*, ed. J. M. Robson. London: Routledge.
- Moore, G. E. (1901–02). "Truth and Falsity," in J. Baldwin (ed.), *Dictionary of Philosophy and Psychology*. London: Macmillan, 716–18.
- Moore, G. E. (1953) [1910–11]. *Some Main Problems of Philosophy*. London: George Allen & Unwin.
- Neale, S. (2001). *Facing Facts*. Oxford: Clarendon Press.
- Pitcher, G. (1964). "Introduction," in G. Pitcher (ed.), *Truth*. Englewood Cliffs, NJ: Prentice Hall, 1–15.

- Russell, B. (1903). *The Principles of Mathematics*. Cambridge: Cambridge University Press.
- Russell, B. (1994) [1905]. "The Nature of Truth," in *The Collected Works of Bertrand Russell*, vol. 4, ed. A. Urquhart. London: Routledge, 492–506.
- Russell, B. (1971) [1912]. *Problems of Philosophy*; reprinted at Oxford: Oxford University Press.
- Russell, B. (1984) [1913]. *Theory of Knowledge: The 1913 Manuscript*. London: George Allen & Unwin.
- Russell, B. (1956) [1918]. "The Philosophy of Logical Atomism," in *Logic and Knowledge: Essays 1901–1950*, ed. R. C. Marsh. London: George Allen & Unwin, 177–281.
- Russell, B. (1956a) [1919]. "On Propositions"; reprinted in *Logic and Knowledge: Essays 1901–1950*, ed. R. C. Marsh. London: George Allen & Unwin, 283–320.
- Salmon, N. (1986). *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Soames, S. (1987). "Direct Reference, Propositional Attitudes, and Semantic Content," *Philosophical Topics* 15: 47–87.
- Tarski, A. (1983) [1935]. "The Concept of Truth in Formalized Languages," in his *Logic, Semantics, Metamathematics*, ed. J. Corcoran. Indianapolis: Hackett, 152–278.
- Wettstein, H. (1986). "Does Semantics Rest on a Mistake?" *The Journal of Philosophy* 83: 185–209.
- Wittgenstein, L. (1995) [1919]. Letter to Russell: 19.8.1919, in B. McGuinness and G. H. von Wright (eds.), *Ludwig Wittgenstein: Cambridge Letters*. Oxford: Blackwell, 124–6.
- Wittgenstein, L. (1961) [1921]. *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness. London: Routledge. (1922 translation by C. K. Ogden.)
- Wittgenstein, L. (1961a) [1913]. "Notes on Logic," in *Notebooks 1914–1916*, ed. G. H. von Wright and G. E. Anscombe. Oxford: Blackwell, 93–106.

CHAPTER 10

THE IDENTITY THEORY OF TRUTH

STEWART CANDLISH AND NIC DAMNJANOVIC

10.1 BOTH PLATITUDINOUS AND IMPLAUSIBLE?

WHEN one thinks truly, what one thinks is what is the case. That is, when one thinks truly that wombats are fast runners, then what one thinks—*that wombats are fast runners*—is the same as what is the case—*that wombats are fast runners*.¹ In this truistic thought lies the germ of the identity theory of truth. For, in the broadest terms, the identity theory of truth holds that truth is a matter of identity between how things are and how one takes them to be.

So characterized, however, the identity theory can seem *merely* truistic. More substantive versions of the theory are phrased in terms of particular types of entity such as truth-bearers and truth-makers, or perhaps true propositions and facts. In particular, identity theorists, motivated in part by the above truism and in part by dissatisfaction with the correspondence theory of truth, typically assert that when a bearer of truth and falsity, such as a proposition, is true, there is no “ontological gap” between it and the way things are.

Yet it is precisely the absence of an ontological gap between truth-bearer and truth-maker that can make the identity theory seem implausible. For a start, anyone who holds that truth-bearers are linguistic entities will find the identity theory a non-starter. For the world is certainly nothing like Figure 10.1 (overleaf).

But things seem hardly better if we treat propositions as the primary bearers of truth. Propositions are usually thought to be abstract objects; reality is not populated

¹ This way of putting the truism—though not the particular example—comes from John McDowell (1994b: 27).

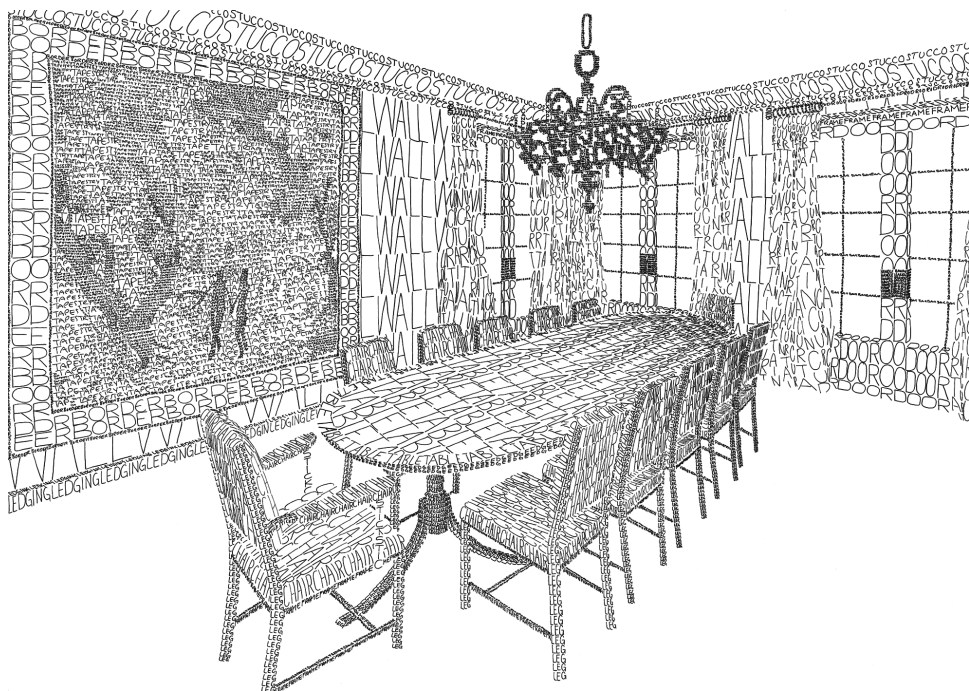


FIGURE 10.1 Thomas Broomé: *Modern Mantra* — Dining Room

(solely) with abstract objects. Propositions, presumably, have conceptual structure; but why think reality is (solely) structured according to our conceptual divisions? On a standard conception of propositions, they are composed of concepts; but reality is not (solely) composed of concepts, and facts, in particular, are presumably composed of objects, properties, and relations. Our platitude seems to have become rather less platitudinous.

Partly for this reason, the identity theory of truth remained less a theory than a subterranean influence throughout most of the twentieth century. Although it was explicitly endorsed by some of the British idealists, and also briefly by Bertrand Russell in the early 1900s, it wasn't until the 1990s that it re-emerged as a generally acknowledged theory of truth.² Since then it has been endorsed by a growing number of philosophers.

Even before the revival of the identity theory of truth, however, many argued that facts and true propositions are identical. Since this identity claim is the most familiar version of the identity theory, we will focus on the plausibility of that identification for much of what follows. It is important to keep in mind, however, that identifying facts with true propositions does not on its own guarantee an identity theory of truth. A theory of

² Two early papers in the re-emergence were Cartwright (1987) and Candlish (1989). The latter paper is responsible for the appearance in the literature of the label the "identity theory of truth."

truth, at least in the sense in which the identity theory is meant to compete with correspondence theories, coherence theories, and so on, is a theory about the nature of truth. To count as such, we must say more than that true truth-bearers are numerically identical with truth-makers: we must add the claim that the nature of truth is found in this identity.

10.2 TRUE PROPOSITIONS AND FACTS

Many philosophers have held that it is simply part of our ordinary conception of facts that they are identical with true propositions. Here is Frege, for example: “Facts, facts, facts’ cries the scientist if he wants to emphasize the necessity of a firm foundation for science. What is a fact? A fact is a thought that is true.”³ Even apart from the self-styled identity theorists to be discussed below, philosophers such as Bertrand Russell, C. J. Ducasse, Rudolf Carnap, Frederic Fitch, Arthur Prior, A. D. Woozley, P. F. Strawson, Timothy Williamson, and Richard Gaskin have agreed.⁴ On the other hand, many philosophers have held that on our ordinary conception of facts they are clearly distinct from propositions.⁵ In this section we consider some general arguments that attempt to take us beyond the deadlock of intuitions.

10.2.1 Arguments for the identification

One argument for the identity of facts and true propositions can be drawn from Frege himself:

A correspondence, moreover, can only be perfect if the corresponding things coincide and so are just not different things . . . It would only be possible to compare an idea with a thing if the thing were an idea too. And then, if the first did correspond perfectly with the second, they would coincide. But this is not at all what people intend when they define truth as the correspondence of an idea with something real. For in this case it is essential precisely that the reality shall be distinct from the idea. But then there can be no complete correspondence, no complete truth. So nothing at all would be true; for what is only half true is untrue. Truth does not admit of more and less.⁶

³ Frege (1918: 25; translation slightly emended).

⁴ Russell (1937 [1903]); Ducasse (1940); Carnap (1947: 28); Woozley (1949: 169–71); Fitch (1971: 100); Prior (1971: 5); Strawson (1998: 403); Williamson (2000: 43), and Gaskin (2008: 58).

⁵ It may be that the disagreement is largely spurious, and that there are two different conceptions of facts—one that treats facts as propositional, the other as worldly. For a discussion of this thought in relation to theories of truth, see Shorter (1962).

⁶ Frege (1918: 3).

For Frege, perfect correspondence requires complete identity between truth-bearer and truth-maker, and thus so does truth.⁷ In other words, truth *requires* identity: anything less would be falling short of capturing things as they really are.⁸ Here we have a general motivation not just for identifying facts with true propositions, but for the identity theory of truth. Yet while there is surely something to Frege's thought, and many have felt its pull, the correspondence theorist has a quick reply: truth may require identity, but only identity in some respect, such as sameness of logical form.⁹

A better approach might be to develop the platitude with which we began. That is, we might try to argue from McDowell's truism,

When one thinks truly, what one thinks is what is the case,

to the identity of facts and true propositions. The natural way to do this is to add the following two thoughts:

When one thinks truly that p, what one thinks is the proposition that p.

When one thinks truly that p, what is the case is the fact that p.

Substituting the terms "the proposition that p" for "what one thinks," and "the fact that p" for "what is the case" in McDowell's original truism we get:

When one thinks truly that p, the proposition that p is the fact that p.

And this is the identity claim we are after. But what should we make of this argument?¹⁰

Our first additional premise relies on a propositionalist analysis of *that*-clauses: to say that someone thinks that p is to say that they stand in the relation of thinking to the proposition that p. This analysis is controversial,¹¹ but it is much less so amongst those who are willing to accept the existence of propositions. Likewise, the second additional premise relies on treating both "what is the case" and "the fact that p" as singular terms.

⁷ For this sort of reason one can see the identity theory of truth as a limiting case of the correspondence theory. Chisholm (1966: 88) and Lewis (2001: 277) do, for example. But most identity theorists of truth do not.

⁸ Compare also Mackie (1973: 57): "If the best we could achieve was that our statements should somehow correspond to what is there, we should still be falling short of having things *just as we state them to be*" (emphasis added).

⁹ Although Frege had an argument against this response—his famous regress argument—few have been convinced. For discussion, see Kalderon (1999).

¹⁰ For a related argument, see David (2001: 686). David's argument does not seem to get us all the way to the identity theory since its conclusion is "For every *x*, *x* is a true proposition iff *x* is a fact." This isn't quite right since it doesn't specify which fact the proposition is to be identical with. Beall (2000) has argued that there is no problem of specification, but we are unconvinced by his dissolution of the problem.

¹¹ Some disjunctivists might give a propositionalist analysis for attributions of false beliefs, but say that when one believes truly what one believes is *the fact that p* (where facts are distinct from true propositions). This gives us an even quicker argument for an identity of the objects of belief and facts.

But, for those who are happy with talk of facts as entities, this should not be a problem: the fact that *p* is one of the things that is the case; the conjunction of all facts is how things are.

A more likely source of concern is the logical form the argument attributes to the truism. It might be thought that it should not be read as identifying (seemingly) two entities. Instead, we could read the truism as roughly equivalent to the following:

When one thinks truly that *p*, it is the case that *p*.

Now, it is plausible that “it is the case that *p*” means, roughly, “*p*.” Our truism, therefore, would amount to little more than the familiar equivalence schema:

It is true that *p* if and only if *p*.

And, from the equivalence schema, surely no substantive identity claim follows.

So, whether there is a good argument for identifying facts and true propositions here depends on difficult questions about the appropriate reading of the original truism. However, the case for the facts/propositions identity can draw support from a reflection on the meaning of *that*-clauses. It is a common assumption that, in “factive” contexts at least, *that*-clauses refer to facts, just as, in other contexts, they refer to propositions. There are various tests for factive contexts, but the basic idea is that in such contexts the sentences embedded must be (or are presupposed to be) true, and this requirement holds even if the embedding context is internally negated.¹² For example, in both the following two sentences, “she openly mistreated her son” is presupposed to be true.

That she openly mistreated her son scared me.

That she openly mistreated her son did not scare me.

In such contexts, it is plausible that the *that*-clause refers to facts. We can, after all, typically insert “the fact” before the *that*-clause without change of meaning. For example:

The fact that she openly mistreated her son scared me.

Supposing that *that*-clauses sometimes refer to facts, it may appear that they are therefore systematically ambiguous.¹³ Yet there are also cases in which *that*-clauses are most naturally read as being univocal. For example:

John remembered that the sun is a star and Susie came to believe that the sun is a star.

John saw that the lamp was broken and it was true that the lamp was broken.

¹² See e.g. Vendler (1975) and Parsons (1993).

¹³ David (2001) and Künne (2003) both argue, against the identity theory, that *that*-clauses are systematically ambiguous.

If we read the *that*-clauses above as univocal, we have three options: either all refer to facts and not propositions, or all refer to propositions and not facts, or facts and true propositions are identical. The first option has the unwelcome consequence that facts are capable of being true or false: cf. “It was true that the lamp was broken.” The second implies we can see true propositions: cf. “John saw that the lamp was broken.” Identifying true propositions with facts provides a simple way around these problems: although we still are said to see true propositions, we should be comfortable with this if it means no more than that we see facts; and although facts are still said to be true this is not in addition to propositions being true or false, since facts just are true propositions. On the simplest semantic treatment of *that*-clauses, then, they are univocal and refer always to propositions, some of which are also facts.¹⁴

10.2.2 Arguments against the identification

Yet while there might be a linguistic argument for identifying true propositions and facts, there are also linguistic arguments against the identification. For example, Harman asks us to compare

The fact that there was an explosion in the basement caused the fire

with

The true proposition that there was an explosion in the basement caused the fire.¹⁵

Even when the former is true, the latter seems false. Likewise, compare

The fact that she openly mistreated her son scared me

with

The true proposition that she openly mistreated her son scared me.

While these and other examples raise suspicion, however, they should not be taken as conclusive. For perhaps if one gets the metaphysics of propositions right it will become apparent how propositions can cause things and scare people, so that our intuitions about these cases fade. We will discuss the relevant metaphysics of propositions below.

¹⁴ Parsons (1993) and Moffett (2003) both point out that treating facts as true propositions would be the neatest way to deal with the complications factive contexts create. However, both reject the identification, though for different reasons.

¹⁵ Harman (2003). For further discussion see Dodd (2000: 90–3). Slote (1974: 99) expresses similar concerns to Harman.

There is another, more often cited, problem for the identification of facts and true propositions that appeals directly to the metaphysics of propositions and facts, however. This argument goes back at least to Moore.¹⁶ It is this. Propositions exist necessarily, or in all possible worlds. Some facts, however, such as the fact that Australia is 18 per cent desert, do not exist in every possible world, since in some worlds Australia receives a uniformly high rainfall. Therefore the proposition that Australia is 18 per cent desert and the fact that Australia is 18 per cent desert are distinct entities.

This standard argument has a standard response due to Richard Cartwright.¹⁷ In a world in which it is not the case that Australia is 18 per cent desert, the entity which in our world is the fact that Australia is 18 per cent desert still exists—*viz.* the proposition that Australia is 18 per cent desert—only in that world it does not have the property of being a fact. That is, there is one entity—the proposition that Australia is 18 per cent desert—which is true in some worlds and false in others. And in every world in which it has the property of being true, it also has the property of being a fact. According to Cartwright, this is much like saying that the author of *Word and Object* exists even in worlds in which neither he nor anyone else composed *Word and Object*.

The standard response has a drawback, however, as it implies that facts are necessary existents. Thus, it is committed to the view that even in the possible world in which Australia has a uniformly high rainfall, *there is* the fact that Australia is 18 per cent desert, or, alternatively, the fact that Australia is 18 per cent desert exists. This seems rather different from saying that even in a world in which no one composed *Word and Object*, the author of that book may exist. For the reading we can give of the latter is that the actual author of *Word and Object* exists in worlds in which he didn't compose that book. But it seems at least odd, if not in fact false, to say that the actual fact that Australia is 18 per cent desert would exist even if Australia had uniformly high rainfall. For this reason, Kit Fine describes the position Cartwright tries to exploit as “definitely wrong” (1982: 50).

While Fine is surely right that, at least on the most familiar philosophical accounts of facts, it is natural to say that facts exist in some worlds but not others, it is plausible that we also have a conception of facts that fits more neatly with Cartwright's response. On this alternative conception, being a fact is a property that some things can have or lack contingently. Here is some evidence that we have such a conception:

You believe *that Mr X's papers are profound*, but that is not a fact. However, if it were a fact, that would be enough to make them interesting papers.¹⁸

In this example, it appears that we use a singular term “that Mr. X's papers are profound” to refer to some entity which may have or lack the property of being a fact. If this is right, then, just as propositions have two poles, true and false, so there are entities

¹⁶ Moore (1953: 308). (Note that these lectures of 1910–11 long antedate their publication.) The argument is endorsed by Ayer (1971: 211) and Fine (1982: 46–7). See also Slote (1974: 98–9).

¹⁷ Cartwright (1987: 77–8). The response is endorsed by Dodd (2001: 87) and David (2001: 688–9).

¹⁸ This example is adapted from McGrath (2003: 162).

which are bipolar in that they can either be facts or not. To identify facts and true propositions is just to say that it is *propositions* that can be facts or non-facts (as well as being true or false).¹⁹

So the situation is this. There are some general reasons to posit the identity of facts and true propositions. And it is at least arguable that we have a conception of facthood as one of two poles an entity may have or lack and according to which conception, therefore, it is at least not obviously incoherent to identify facts and true propositions.

10.3 THE METAPHYSICS OF FACTS AND PROPOSITIONS

Yet we still have the problem that propositions and facts seem to be things with very different natures: how could they be identical? In this section we will consider some possible ways to make the identification coherent and consider whether they are compatible with an identity theory of truth.

10.3.1 Fregean facts

In Frege's mature philosophy of language he distinguished the realm of reference from the realm of sense. The realm of reference consists of the things to which we refer, including ordinary objects as the referents of proper names. The realm of sense consists of what Frege called the modes of presentations of objects. Thus a proper name has a sense which is a mode of presentation of, or perhaps a way of thinking of, its referent. Likewise, a sentence expresses a sense, called a Thought, which is a mode of presentation of the referent of a sentence, that referent being, for Frege, one of the two truth-values, the True or the False. A proposition, then, is a Thought, which is a compound sense composed of other senses. And we have already seen that, for Frege, a "fact is a thought that is true."²⁰ Thus facts must also be composed of senses, belonging to the realm of sense, not reference;²¹ and they are abstract entities, the contents of thoughts and the sense of sentences.²²

While such a view makes an identification of facts and propositions coherent, it is not obvious whether identifying facts with Fregean propositions is consistent with our ordinary conception(s) of facts, or with the theoretical roles that we would like facts

¹⁹ For a discussion of the different conceptions of facts, as unipolar or bipolar, in relation to the identity theory, see Engel (2001).

²⁰ Frege (1918: 25).

²¹ See Dummett (1981: 177).

²² While Frege identified facts and true propositions, he also held that truth is a primitive, indefinable property. He was led to this view by his already-mentioned regress argument.

to play. One reason for concern might be that Fregean propositions are individuated too finely to be facts.²³ Since Fregean propositions are composed of modes of presentation, the proposition that Cicero is talking is different from the proposition that Tully is talking: but, it might be thought, the fact that Cicero is talking is no different from the fact that Tully is talking. It is not clear that this is so, however. Consider:

Samantha acknowledged the fact that Cicero is a famous orator.

Samantha acknowledged the fact that Tully is a famous orator.

As it is plausible that the first might be true without the second's being true, it is arguable that facts should be individuated as finely as Fregean propositions.

The greater cause for concern is that Fregean propositions are not "worldly" items in two senses: they are abstract rather than concrete; and they are composed of abstract entities rather than the denizens of the spatio-temporal universe. Yet, for many philosophers, facts are worldly in both senses, or nothing at all. For them, it is the role of facts to serve as the worldly entities which account for, or ground, the truth of propositions. Moreover, identity theorists can't—in the hope of deflecting this concern—endorse the more usual conception of facts and yet offer the Fregean conception as another, equally legitimate, conception of facts. Since truth surely depends on being, or on how things are with the world, as long as worldly facts are admitted it will seem as if truth is the relation that holds between propositions and *those* facts.²⁴ So an identity theory of truth that treats facts as Fregean will need to remove the temptation to posit worldly facts entirely.

The classic attack on worldly facts comes from Strawson (1950). Strawson argued not only that facts are individuated as finely as propositions, but that the only way we *can* individuate them is by using the same *that*-clauses that refer to the propositions which they are meant to make true. The only way to refer to the fact that Cicero is a famous orator is to use the same *that*-clause that we use to refer to the proposition that Cicero is a famous orator. With only this way of referring to them, we have no conception of facts independent of propositions. In Armstrong's memorable phrase, for Strawson facts are but the "tautological accusatives" of true propositions.²⁵

This critique of worldly facts takes us only so far, though. To overcome it, all a friend of worldly facts must do is provide the missing independent conception of them, perhaps by describing them as entities constructed out of worldly objects. For this reason, Strawson's main criticism of worldly facts was that they serve no useful purpose. To look for worldly facts is in fact to commit a sort of logical, or category, mistake.²⁶ One source

²³ David (2001: 692).

²⁴ Well, the truth of at least some propositions, such as the proposition that Cicero is a famous orator, supervenes on being.

²⁵ Armstrong (1997: 19).

²⁶ Compare Harman (2003) who thinks the category mistake is to *identify* true propositions and facts.

of the category mistake is the confusion of stating and referring: a proposition is not used to refer to a thing, but to state something. That is, while propositions are about things in the world (people, objects, etc.), “there is nothing else in the world for the statement itself to be related to either in some further way of its own or in either of the different ways in which [the] different parts of the statement are related to what the statement is about” (Strawson 1950: 166). It is the demand for such an entity which is the category mistake, and “the demand for something in the world *which makes the statement true . . . or to which the statement corresponds when it is true*, is just this demand” (ibid., Strawson’s emphasis). With the category mistake pointed out, presumably, the temptation to posit worldly facts should evaporate like dew in the morning sun.

But it does not. Even if we agree that propositions don’t refer to facts, the intellectual urge remains to investigate the relations between true propositions and reality. And one approach to satisfying this urge is to posit entities which make propositions true: that is, truth-makers. Julian Dodd, the most prominent contemporary defender of a Fregean style identity theory of truth, has recognized this, and argues instead that this truth-making approach rests on the following principle, which lacks sufficient motivation:

TM. For the proposition that *p* to be true, there must exist at least one entity, distinct from the proposition that *p*, whose existence entails that the proposition that *p* is true.²⁷

Dodd’s main target, then, is the truth-maker project: the idea that we must hypostatize ways the world is into entities that are distinct from propositions and which make some of those propositions true.

While we can’t consider the arguments for and against the truth-maker project here, suffice it to say that even if Dodd is right, a question remains: namely, “What is the relation between true propositions and reality?” Since truth is in some sense grounded in being, without an answer to that question, an identity theory of truth like Dodd’s seems to evade the most fundamental question of all: since the only account of facts it provides us with is that they are true Fregean propositions, it gives us no account of the nature of truth.²⁸ Instead, it looks like an attack on the truth-maker project combined with the theory that facts are nothing but true propositions.

Dodd almost admits as much himself. The goal of his identity theory, he says, is not to define truth, but to point up the error in the correspondence theory.²⁹ Tellingly, he supplements his identity theory with a deflationary account of truth. More specifically, he provides a Horwich-inspired minimalist account of the relations between true propositions and reality.³⁰ That is, the equivalence schema tells us all there is to know about how truth is grounded in being:

ES. The proposition that *p* is true if and only if *p*.

²⁷ Dodd (2000: 9).

²⁸ Candlish (1999b: 235–6) makes this complaint.

²⁹ Dodd (2000: 124–8).

³⁰ Dodd (2000: ch. 6); Horwich (1998).

Whether one is sympathetic to deflationism or not, it is clear that Dodd needs to appeal to some story, other than the standard correspondence account, about the relation of true propositions to reality to provide a theory of truth.³¹ Once he does so, however, it appears that the real work of the theory of truth is being done by that story, and not the identity theory. So an identity theory of truth that identifies facts with true Fregean propositions may be coherent, but without supplementation should not be thought of as a theory of truth. And yet the necessary supplementation appears to be the real theory of the nature of truth.

10.3.2 Russellian propositions

The most familiar alternative to the Fregean account of propositions is the “Russellian” account. Russellian propositions, like Fregean propositions, are structured entities. Unlike Fregean propositions they have real-world objects, properties, and relations as their constituents. Russell, of course, held this view, as did Moore.³² More recently it has been defended by Nathan Salmon and Scott Soames, among others.³³

At first thought, true Russellian propositions are very well suited to be identified with facts.³⁴ The dominant conception of facts is that they are worldly entities composed of real-world objects, properties, and relations. Worldly facts and Russellian propositions thus have the same constituents. For this reason Russellian propositions are also presumably as well placed as worldly facts to enter into causal relations (*contra* Harman’s worry mentioned in section 10.2.2). Since Russellian propositions have real-world constituents, they may also be held to exist contingently: in worlds in which one or more of the constituents don’t exist, it might be thought, the proposition does not exist either. Thus this sort of proposition *may* be immune from the other worry, discussed above, that the modal properties of facts and propositions are different.³⁵

³¹ Dodd’s theory may even be compatible with a theory of truth-making. That is, it could be if McGrath’s (2003) deflationary theory of truth-making works, according to which the truth-maker for the proposition that *p*, is that *p*.

³² It is important to note that Moore did not himself unequivocally identify facts and true propositions. His early papers (Moore 1899; 1901) are suggestive, but he later (Moore 1953: 260–2) describes his earlier view as the position that facts are identical with the truth of propositions, or the being true of a proposition. As Fine says, the being true of a proposition is no more a true proposition than the wisdom of a man is a wise man. (His earlier papers speak of “truths” being no different from reality, but “truths” is ambiguous between true propositions and the truth of a proposition.) Moore’s position is an interesting cousin of the identity theory, but we cannot discuss it here. See Slote (1974) and Fine (1982) for defenses of it.

³³ Salmon (1986; 1989a; 1989b); Soames (1985; 1987; 1989).

³⁴ What about true *negative* propositions? Our view is that an identity theory faces no more (and no less) difficulty than the correspondence theory on this matter, and we shall accordingly ignore it.

³⁵ There are other problems that follow from treating propositions as contingently existing. For a brief and helpful discussion with further references see Fitch and Nelson (2007: section 5).

Yet it is precisely the closeness of Russellian propositions to facts that creates difficulties. Just as there is a suspicion that Fregean propositions are too fine-grained to serve as facts, so there is a suspicion that fact-like entities are too coarse-grained to serve as propositions. For example, the following two sentences presumably express the same Russellian proposition:

Cicero is Tully.

Cicero is Cicero.

Likewise, if one adopts an analysis of belief sentences whereby they express the proposition that a person stands in the belief relation to the proposition expressed by the that-clause, then the following two sentences express the same proposition:

Caesar believes that Cicero is Cicero.

Caesar believes that Cicero is Tully.

Yet it seems as if the first two sentences differ in the information they convey and the second two could differ in truth-value. Such consequences are usually taken as good evidence that sentences express different propositions. Nevertheless, Soames and Salmon have vigorously defended their views against such criticisms.

A deeper worry emerges, however, if we think more about the nature of propositions. Russellian propositions are usually thought of as entities in some sense “over and above” their constituents. They are not simply the collection of their constituents, since different propositions can have the same constituents—witness the proposition that Caesar pardoned Cicero and the proposition that Cicero pardoned Caesar, both of which on a Russellian account contain Cicero, Caesar, and the pardoning relation. Moreover, a proposition must be a unity in the sense that the constituents do not just sit alongside each other, like peas in a pod, but are bound together in such a way that they *say something*, or *have truth-conditions*. So any theory of structured propositions must answer the question, “What binds the constituents of a proposition together so that they make an appropriate, unified whole?” (Not to mention the further question brought out by our example, “Why this whole rather than that?”) And this question poses a dilemma for those who would use a Russellian account of propositions as part of an identity theory of truth.

On one horn of the dilemma, we can answer the question in such a way that the propositions can be identical with worldly facts. That is, we can follow Russell (1937 [1903]) in holding that what unifies the constituents Cicero, Caesar, and the pardoning relation, in the proposition that Caesar pardoned Cicero, is the pardoning relation itself. When the proposition is true, therefore, it just is the fact that Caesar pardoned Cicero: it contains the same constituents unified in the same way. The problem is that there is now no room for falsity. For if the pardoning relation is the glue that binds Cicero and Caesar

together in the proposition that Caesar pardoned Cicero, then, even if the proposition is false, it is still identical with the fact that Caesar pardoned Cicero and so the fact that Caesar pardoned Cicero must exist. This problem is worrying enough on its own, but it is especially worrying for someone who wants to hold an identity theory of truth. This is because it is not a unique feature of *true* propositions that they are identical with a fact—false propositions are too! So an identity theory of truth is simply unavailable.³⁶ Russell himself adopted primitivism about truth—the view that truth is a *sui generis*, unanalyzable property.³⁷

On the other horn, we make room for false propositions by finding an account of the unity of a proposition according to which whatever unifies the proposition that *p* is not what also unifies the fact that *p*.³⁸ This allows the proposition that Caesar pardoned Cicero to be a unity without its being the case that Caesar pardoned Cicero. If a workable account of this sort can be found, the Russellian view of propositions may be saved. And yet the identity theory of truth is still lost. For now a proposition cannot be identical with the relevant fact since they are unified in different ways.

The identity theory of truth cannot be defended on either horn of the dilemma.

10.3.3 States of affairs and possibilities

There is, however, a way to avoid the two horns. The first assumes that what unifies a fact is also what is responsible for its obtaining. The relation of pardoning, say, both unifies the fact that Caesar pardoned Cicero and is responsible for its being the case, being a fact, that Caesar pardoned Cicero. We could instead, though, say that there is one account of what unifies a fact and a distinct account of what makes it obtain. This option is more clearly put if we adopt the terminology of states of affairs. It is one thing for a state of affairs, such as that Caesar pardoned Cicero, to be unified, and so to exist, or have being; it is another for that state of affairs to obtain, or be a fact, so that Caesar pardoned Cicero. The state of affairs exists whether or not it obtains, whether or not Caesar pardoned Cicero, but only obtaining states of affairs are facts.³⁹

States of affairs are thus bipolar, like propositions, and are well suited to be identical with them. Moreover, we can identify propositions and states of affairs without running into the problems Russell's account faced. A false proposition is identical with

³⁶ Unavailable, that is, to anyone not prepared to countenance *trivialism*, the doctrine that every statement is true. It is understandable that Graham Priest, a dialetheist, has found himself obliged to discuss trivialism (Priest 2000). He argues that it cannot be believed; this, of course, is consistent with its being true.

³⁷ Russell (1937 [1903]: section 52).

³⁸ For a recent account that fits this description see King (2007).

³⁹ Thus this account of states of affairs is incompatible with the view, called "hard actualism" by Adams (1979), that facts are the only states of affairs.

a non-obtaining state of affairs and so the mere existence of the proposition that Caesar pardoned Cicero does not entail that it is a fact that Caesar pardoned Cicero. This means that there is also a distinction between true and false propositions. Where a true proposition is identical with a fact—an obtaining state of affairs—a false proposition is not.

In broad outline, but with a different terminology, the view just described is that of Meinong (1983 [1902]). More recently, it has also been defended by Chisholm (1966) and Gaskin (2008). While Meinong and Gaskin offer this account as part of a broader theory of intentionality or reference, Chisholm explicitly offers it as part of a theory of truth (though he does not use the label, not then invented, “identity theory of truth”).

There are also other, similar, ways to understand states of affairs, according to which they cannot be identified with *Russellian* propositions, but may be identified with propositions differently conceived. On one such understanding, states of affairs are not necessarily made of real-world constituents, but instead are abstract entities that exist necessarily. According to some versions of this view, states of affairs are possible worlds. And possible worlds, or sets of possible worlds, have many of the right features to be both states of affairs and propositions. Robert Stalnaker (1976), for example, defends the view that propositions are sets of possible worlds. And, as Baldwin points out, false Russellian propositions have very much the same status as non-actual worlds in David Lewis’s (1986) theory of possible worlds.⁴⁰

On Lewis’s theory, all possible worlds are concrete entities—even the *merely* possible ones. Amongst the set of possible worlds the actual world is (is identical with) the real world. This is in contrast to views according to which possible worlds are mere representations or “ersatz” worlds. On those views, the actual world cannot be the real world, since possible worlds are all representations and the real world is not a representation. The actual world is the world that *represents* the real world. But for Lewis, the actual world just *is* the real world. Now, if propositions are sets of possible worlds, then a proposition is presumably true just in case the actual world is a member of it. This does not quite give us an identity between propositions and the real world, but it gets us something very close, and very far from, say, correspondence. Moreover, a person who believes a maximally determinate true proposition stands in a very direct relation to reality: they stand in the belief relation to a set whose sole member is the real world.

If any of these accounts of propositions and states of affairs, possible facts, or possible worlds can be defended, then facts and true propositions may be identical (or something close to it). If they are, what does this tell us about the nature of truth? On these accounts, all propositions, whether true or false, are identical with states of affairs. So what is special about *true* propositions is that they are identical with states of affairs which obtain (or possibilities that are actual, or . . .), i.e. facts. If this is to illuminate the nature of truth, then presumably the idea is that what it is for a proposition to be true is for it to be a state of affairs that obtains. In other words, the property of truth just is the

⁴⁰ Baldwin (1991: 46, 50–1); see also Candlish (1999a: 218–20).

property of obtaining. But what is this latter property? It is hard to imagine that there could be a substantive theory about it: it seems to be a primitive. If so, then truth is a primitive property too.⁴¹ On the other hand, there may be a sort of deflationary theory of the property given by the instances of the following schema:

The state of affairs that *p* obtains if and only if *p*.

If so, then deflationism seems the right approach to truth. Or perhaps primitivism and deflationism amount to much the same.⁴² Either way, this identity theory of truth appears to be, at root, either a primitivist or deflationist theory of truth.

10.3.4 Disjunctivism

The dilemma we posed for an identity theory of truth that makes use of Russellian propositions made the assumption, which we dropped in section 10.3.3, that what unified a fact was also what was responsible for its obtaining. It also made another assumption, which we upheld in section 10.3.3, but which can be dropped instead: it assumed that true and false propositions are fundamentally the same kind of entity, unified in the same way, which differ simply in possessing either the property of being true or that of being false.

If we were to drop the latter assumption, we could hold a version of disjunctivism about propositions, according to which propositions are not all one kind of thing.⁴³ For example, true propositions might be composed of objects, properties, and relations and might be unified in a way that guarantees they are facts: the proposition that Caesar pardoned Cicero, that is, might be unified by the pardoning relation itself, so that it is no different from the fact that Caesar pardoned Cicero. False propositions, such as the proposition that tigers have quills, might be unified in some other way which ensures that they are not facts (or at least not the fact that tigers have quills). Or, alternatively, perhaps true and false propositions differ in their constituents. However the details are worked out, by allowing two fundamentally different types of proposition, it is possible to identify facts with true propositions without also identifying them with false propositions.

Disjunctivism of this sort would also be compatible with an identity theory of truth. For it is open to a disjunctivist to say that the nature of truth is that a

⁴¹ Meinong explicitly endorsed this view. A state of affairs obtaining is what Meinong calls an objective being factual. And factuality is “a fundamental property which admits of no definition” (quoted by Findlay 1933: 76). See also Gaskin (2008: 125, esp. fn. 255).

⁴² See Sosa (1993) and Stoljar and Damjanovic (2007).

⁴³ The disjunctivist label is meant to reflect a connection to the well-known disjunctivist position in the philosophy of perception. It is perhaps unsurprising that both McDowell and Fish, identity theory sympathizers, are also disjunctivists in the philosophy of perception. See McDowell (1994a) and Fish (2009).

proposition is true when it is a fact. The property of being true is the property of being a fact. Moreover, it is possible that we could give an account of facts independently of true propositions so that this theory had real content. For example, facts could be individuated by their constituents and the way they are unified. A similar story might be told for falsehood.

But while the outlook is optimistic in these directions, disjunctivism has more trouble with the standard argument against identifying facts and true propositions (see section 10.2.2). For on this view facts are not necessarily existing entities: if they were, then the actual fact that Caesar pardoned Cicero would exist in all worlds (though it wouldn't be a fact in all worlds) and so the actual proposition *Caesar pardoned Cicero* would exist in all worlds, since the actual true proposition is identical with the actual fact. However, according to the disjunctivist, the actual proposition *Caesar pardoned Cicero* can't be the proposition *Caesar pardoned Cicero* in all worlds, as false propositions are essentially different from true propositions, and in some worlds it is false that Caesar pardoned Cicero. Thus, if actually true propositions/facts existed in all worlds, then the proposition *Caesar pardoned Cicero* could possibly not be that proposition. But this is absurd—propositions are essentially the propositions they are. Probably, the most plausible position for a disjunctivist to adopt is that there is a proposition *Caesar pardoned Cicero* in all worlds, but different entities play that role in worlds in which it is true that Caesar pardoned Cicero and in worlds in which it is false. Even this position has its costs, however, since, given the propositionalist reading of *that*-clauses, it implies that were it not the case that Caesar pardoned Cicero, I would not be able to stand in the believing relation to the same entity that I do, *viz.* the actual proposition that Caesar pardoned Cicero. For in those circumstances I would stand in the believing relation to some other, essentially distinct, entity.

Disjunctivism is at least not obviously false. But it is undoubtedly counterintuitive and its counterintuitiveness is a measure of the sort of lengths to which one needs to go in order to identify facts and true propositions while also endorsing a coherent, substantive identity theory of truth. It is therefore worth considering some other routes to the identity theory.

10.4 QUIETISTIC THEORIES

Dodd (1995; 2000) distinguishes between what he calls modest identity theories and robust identity theories. Modest identity theories are those that identify facts with true Fregean propositions. Robust identity theories identify true propositions with worldly facts. Thus, in Dodd's terms, the views presented in the last three subsections are robust identity theories and his own, discussed in section 10.3.1, is modest. This classification runs into trouble, however, in dealing with the identity theory espoused by John McDowell (1994b; 2005). For, on the one hand, McDowell claims that facts are true thinkables (conceived on Fregean lines) and, on the other, he claims that the sum

of facts makes up the world. Dodd, unsurprisingly, has argued that McDowell's position is incoherent.⁴⁴

William Fish and Cynthia Macdonald (2007; 2009), in contrast, have recently outlined an interpretation of McDowell's position which they claim shows that Fregean propositions could be identical with "worldly" facts. Their ingenious interpretation supposes that, for McDowell, a proposition is composed of senses, but *senses themselves can be facts*. Thus a proposition is a molecular fact composed of other facts which are also senses. For example, the fact/true content that this tiger is undernourished is constituted by the fact that this object is a tiger and the fact that it is undernourished.⁴⁵ This hybrid position, neither robust nor modest, allows that facts and Fregean propositions can have the same constituents, while facts are nevertheless "worldly." There is no space to consider the merits of this view here. Suffice it to say that if true propositions have facts as constituents, then there is a question about the constituents of false propositions. While Fish and Macdonald make no explicit claims about false propositions, it is clear that they must adopt one of the positions sketched in the last few sections: either there are no false propositions, false propositions are non-obtaining states of affairs (and facts obtaining states of affairs), or true and false propositions are radically different in kind. For the reasons given above, only the last appears compatible with an identity theory of truth.⁴⁶

Noting McDowell's assertion that his view of truth is meant to be truistic, we might be suspicious of Fish and Macdonald's attribution to him of substantive views about the nature of propositions and facts.⁴⁷ An alternative way of reading McDowell's remarks on truth would be to emphasize the similarities between his views and Jennifer Hornsby's, and so the quietistic elements of his thought. On this approach, the identity *conception* (not *theory*) of truth is seen, not as a contribution to "constructive philosophy," but as an attempt to remove the intellectual pressures which drive us to produce constructive accounts. In one sense the resulting view is substantive, since it is incompatible with substantive philosophical theories (such as the correspondence theory); but in another it is modest, since it confines itself to certain truisms.

The truism McDowell and Hornsby most emphasize is the truism with which we began:

When one thinks truly what one thinks is the case.

So, if one thinks truly that Caesar pardoned Cicero, then what one thinks, that Caesar pardoned Cicero, is what is the case, that Caesar pardoned Cicero. We have already seen

⁴⁴ Dodd (1995; 2000). He retracts this charge in Dodd (2008).

⁴⁵ Fish and MacDonald (2007: 39).

⁴⁶ Given McDowell's (1994a) endorsement of some kind of disjunctivist position in the philosophy of perception, and given that he endorses that view precisely to avoid any intermediaries or "veil of ideas" between true perception and the world, the disjunctivist position described above would appear to be the most acceptable to McDowell himself.

⁴⁷ Dodd (2008) says likewise.

that parlaying this truism into the claim that facts are identical with true propositions is not entirely straightforward. But this difficulty is unlikely to bother McDowell and Hornsby, who prefer to stay closer to the truisms. For example, they tend to avoid talk of propositions and facts and speak instead of thinkables and what is the case. Part of the reason for this terminological choice seems to be to avoid taking propositions and facts as objects or entities that are being identified: Hornsby in particular advises us that the identity thesis should not be thought of as “upping the ontological stakes (so that thinkables are to be treated as OBJECTS).”⁴⁸ In other words, reifying what one thinks and what is the case into truth-bearers and truth-makers moves us beyond the truisms in a way which only encourages the sort of theorizing that pushes us toward the correspondence theory of truth.

Other apparently substantive consequences of McDowell’s and Hornsby’s views might also be illusory. As mentioned above, it follows from the identification of true thinkables (which McDowell and Hornsby treat on roughly Fregean lines) with facts that facts cannot be composed of objects and properties, but must have senses as constituents. This would be harmless if we are willing to see facts as Fregean facts, but McDowell and Hornsby insist that facts are the elements of the world. Is the world therefore composed of senses? In a sense, yes. As McDowell says, it is a natural way of talking to say that the world is all that is the case. And things that are the case, such as Caesar’s pardoning Cicero, are things that we can think. So the world is made up of things we can think. But this remains largely truistic. Furthermore, there is another sense of “the world” in which it is not made up of senses. For one thing, if we resist reifying thinkables and facts, then to speak literally of “constituents” of either is misguided. For another, we can also think of reality as the realm of reference, as things we think *about* rather than things we think, and in this sense of reality it is not made up of senses but of things and properties.

This dualism of sense and reference, however, can make us wonder how these two realms are connected. A now familiar thought is that they are connected because the realm of reference is what makes true thinkables true. And this can lead us to the idea that facts must be part of the realm of reference and made up of more basic elements such as objects and properties. As with other identity theorists, McDowell and Hornsby advise us away from this way of thinking. The realm of reference, in a sense, makes our thoughts true. If it is true that Caesar pardoned Cicero, that is because the object Caesar is related to the object Cicero by the pardoning relation in a way that is, we might say, from left to right. However, this just means that the way things are with those objects determines that it is a fact that Caesar pardoned Cicero. We do not also need to say that the objects and the relation are literally constituents of some entity, the fact. (Thinking of facts as objects tends to lead us astray here.) We thus also do not need to say that the objects and properties are somehow more fundamental than the facts. The realm of sense and the realm of reference are not two sets of things as much as two ways of conceptualizing the one reality.

⁴⁸ Hornsby (1997: 4 fn. 6).

Since their views stay so close to the truisms, it is probably a mistake to think of McDowell or Hornsby as offering a theory of truth. Their goal seems rather to lead us away from substantive theorizing about truth. Specifically, while they undoubtedly hope to convince us that there is “no ontological gap” between mind and world (when we think truly), they do not seem to want to convince us that facts are identical with true propositions, or even that truth-bearers are identical with truth-makers. They are at once more radical and more conservative: they seem to want us to abandon thinking in terms of truth-makers and truth-bearers and remain with more humble truisms about the relations between thought and reality.

As such, the identity *conception*, like other identity *theories*, of truth seems to be compatible with both deflationism and primitivism. In fact, McDowell and Hornsby both point out that their view is distinct from deflationism only if deflationists are committed to denying that the notion of truth can play a central role in theorizing about meaning;⁴⁹ and Hornsby, at least, seems amenable to the idea that truth could be further illuminated, but not defined, by mapping its connections to other semantic concepts by constructing theories of interpretation—but this is just Davidsonian primitivism.⁵⁰

10.5 IDEALISM

Despite his preference for truisms, there is a way in which McDowell sometimes goes beyond them. Sometimes he says not just that when one thinks truly what one thinks is the case, but that the world is within the realm of sense. And this latter way of speaking, though not the former, implies at least a weak form of idealism, as it amounts to the idea that there are no facts that are incapable of being thought: the world is in an important sense “mind-shaped.”

While this idealism might be harmless, or even inessential to McDowell’s project, more robust forms of idealism provide quick ways to secure an identity of thought (or language) and reality. If the world is entirely mental (or linguistic), then there is no problem in supposing that there is no “ontological gap” between mind (language) and world. One clear form of the identity theory, in fact, was endorsed by various nineteenth- and twentieth-century post-Hegelian idealists. Those we have particularly in mind are the monists F. H. Bradley, Harold Joachim, and Brand Blanshard:⁵¹ we shall call them the British idealists.⁵²

⁴⁹ Hornsby (1997: 16–22); McDowell (2005: 87).

⁵⁰ Hornsby (1997: 11–15); Davidson (1996).

⁵¹ Bradley (1911); Joachim (1939 [1906]); Blanshard (1939).

⁵² Candlish (1989) already argues that Bradley was no coherentist; the case is pressed home in detail in Candlish (2007). For a fuller account of the position of Joachim and Blanshard (including a justification for calling Blanshard a “British” idealist), see Candlish and Damnjanovic (2007) and Damnjanovic and Candlish (2013).

The British idealists had a monistic and holistic conception of reality according to which reality is not fundamentally a plurality of self-dependent entities, but a unified whole with each element (if, in thought, we abstract elements from the whole) dependent on all the others. On this view, any thought that abstracts an aspect of reality from the whole cannot represent things as they really are. But such thoughts are not simply false either, since for the British idealists truth and falsity come in degrees. That judgment is the least true which is the most distant from the whole of reality; conversely, the closer a judgment is to the whole, the more true it is. Even so, the truest thought is not simply the one that is maximally determinate—a complete description of the world, as it were. Rather, any thought which divides the world into subject and predicate or objects and relations—in other words any thought which employs the categories of thought—is still an abstraction and so not fully true. The only way a thought can really be adequate to reality is by taking on the very nature of the reality it is meant to be about; and the only way to do that is by *becoming* that reality. In a thoroughgoing monism, thought itself cannot stand outside the all-encompassing whole.

It is because, as monistic idealists, the British idealists also spoke of reality as a coherent whole, and therefore of truth as requiring coherence, that they have often been confusedly thought of as holding a coherence theory of truth. It is crucial to see, however, that the element of coherentism in their views is a result of the more primary identification of true thought with reality. One of the reasons that they should not be thought of as coherence theorists is that their monism rules out dividing up what can be thought into individual thinkables or propositions. Thus they could not accept the coherence theory as typically formulated, which defines truth as a relation amongst propositions. For the same reason, they do not hold an identity theory which asserts that truth is identity between proposition and fact. Nevertheless, they did clearly hold that the nature of truth is fundamentally a matter of identity between thought and reality.⁵³

10.6 CONCLUDING COMMENTS

Construed as a theory of the nature of truth, the identity theory is not easy to defend. The British idealists managed to do so only by abandoning the ideas of distinct truth-bearers and distinct truth-makers and by holding an extreme, idealistic monism. If one wishes to avoid the former commitment, and give an account of truth involving facts

⁵³ Joachim and Blanshard also spoke of the nature of truth being coherence, but it is clear that this claim is a consequence of the more fundamental nature of truth as identity of thought with reality. Bradley, in the end, denied this identity, but as it were from the inside: because reality is more than can be captured by the intellect, the identity required what he called thought's "happy suicide" (Bradley 1897: 152; see also Bradley 1922: 590–1).

and propositions, for example, then it appears that the only option is to opt for disjunctivism and fight a battle that may turn out to be just as difficult to win. Alternatively, one might follow McDowell and Hornsby down the path to quietism. By doing so, one not only gives up the goal of providing a substantive account of the nature of truth, but also may be forced to adopt at least a weak version of idealism.

Perhaps, then, the identity theory of truth should not be treated as an account of the nature of truth. Instead, it may be better thought of simply as an identity theory of two types of entities: namely, truth-bearers and truth-makers, or more specifically, or if one doesn't like the idea of truth-makers, facts and true propositions. This sort of view is compatible with all the major theories of truth—coherence theories, pragmatist theories, primitivism, deflationism, etc.—with the possible exception of the correspondence theory: and it is even possible to see it as a sort of limiting case of correspondence views.⁵⁴ Moreover, even when purged of the ambition to illuminate the nature of truth, defending the identity of facts and true propositions, or truth-bearers and truth-makers, will itself rely on highly controversial theories of at least one of those things.

REFERENCES

- Adams, R. (1979). "Theories of Actuality," in *The Possible and the Actual: Readings in the Metaphysics of Modality*, ed. Michael J. Loux. Ithaca, NY: Cornell University Press, 190–209.
- Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Ayer, A. J. (1971). *Russell and Moore: The Analytical Heritage*. Cambridge, MA: Harvard University Press.
- Baldwin, T. (1991). "The Identity Theory of Truth," *Mind* 100: 35–52.
- Baldwin, T. (ed.) (1993). *G. E. Moore: Selected Writings*. London: Routledge.
- Beall, Jc. (2000). "On the Identity Theory of Truth," *Philosophy* 75: 127–30.
- Blackburn, S. and K. Simmons (eds.) (1999). *Truth*. Oxford: Oxford University Press.
- Blanshard, B. (1939). *The Nature of Thought*. 2 vols. London: George Allen & Unwin.
- Bradley, F. H. (1897). *Appearance and Reality*, 2nd edn. London: Swan Sonnenschein. (Page references here are to the so-called "ninth impression," Oxford: Clarendon Press, 1930.)
- Bradley, F. H. (1911). "On Truth and Copying," reprinted in F. H. Bradley (1914), *Essays on Truth and Reality*. London: Clarendon Press.
- Bradley, F. H. (1922). *The Principles of Logic*, 2nd edn. London: Oxford University Press.
- Candlish, S. (1989). "The Truth about F. H. Bradley," *Mind* 98: 331–48.
- Candlish, S. (1999a). "A Prolegomenon to an Identity Theory of Truth," *Philosophy* 74: 190–220.
- Candlish, S. (1999b). "Identifying the Identity Theory of Truth," *Proceedings of the Aristotelian Society* 99: 233–40.
- Candlish, S. (2007). *The Russell/Bradley Dispute and its Significance for Twentieth-Century Philosophy*. Basingstoke: Palgrave Macmillan.

⁵⁴ See Chisholm (1966: 88); Lewis (2001: 277).

- Candlish, S. and N. Damjanovic (2007). "A Brief History of Truth," in D. Jacquette (ed.), *Handbook of the Philosophy of Science. Volume 5: Philosophy of Logic*. Amsterdam: Elsevier BV, 273–369.
- Carnap, R. (1947). *Meaning and Necessity*. Chicago: University of Chicago Press.
- Cartwright, R. (1987). "A Neglected Theory of Truth," in R. Cartwright, *Philosophical Essays*. Cambridge, MA: MIT Press, 71–93.
- Chisholm, R. M. (1966). *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall.
- Damjanovic, N. and S. Candlish (2013). "The Myth of the Coherence Theory of Truth," in M. Textor (ed.), *Judgement and Truth in Early Analytic Philosophy and Phenomenology*. Basingstoke: Palgrave Macmillan, 157–82.
- David, M. (2001). "Truth as Identity and Truth as Correspondence," in Lynch (ed.) (2001), 683–704.
- Davidson, D. (1996). "The Folly of Trying to Define Truth," *Journal of Philosophy* 93: 263–79.
- Dodd, J. (1995). "McDowell and Identity Theories of Truth," *Analysis* 55: 160–5.
- Dodd, J. (2000). *An Identity Theory of Truth*. New York: St. Martin's Press.
- Dodd, J. (2008). "McDowell's Identity Conception of Truth: a Reply to Fish and Macdonald," *Analysis* 68: 76–85.
- Ducasse, C. J. (1940). "Propositions, Opinions, Sentences, and Facts," *Journal of Philosophy* 37: 701–11.
- Dummett, M. (1981). *The Interpretation of Frege's Philosophy*. London: Duckworth.
- Engel, P. (2001). "The False Modesty of the Identity Theory of Truth," *International Journal of Philosophical Studies* 9: 441–58.
- Findlay, J. N. (1933). *Meinong's Theory of Objects*. Oxford: Clarendon Press.
- Fine, K. (1982). "First-Order Modal Theories III—Facts," *Synthese* 53: 43–122.
- Fish, W. (2009). *Perception, Hallucination, and Illusion*. New York: Oxford University Press.
- Fish, W. and C. Macdonald (2007). "On McDowell's Identity Conception of Truth," *Analysis* 67: 36–41.
- Fish, W. and C. Macdonald (2009). "The Identity Theory of Truth and the Realm of Reference: Where Dodd Goes Wrong," *Analysis* 69: 297–304.
- Fitch, F. (1971). "Propositions as the Only Realities," *American Philosophical Quarterly* 8: 99–103.
- Fitch, G. and M. Nelson (2007). "Singular Propositions," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition), <http://plato.stanford.edu/archives/spr2009/entries/propositions-singular/>
- Frege, G. (1918). "Thoughts," in G. Frege (1977), *Logical Investigations*, ed. P. T. Geach. Oxford: Blackwell, 1–30.
- Gaskin, R. (2008). *The Unity of the Proposition*. Oxford: Oxford University Press.
- Harman, G. (2003). "Category Mistakes in M&E," *Philosophical Perspectives* 17: 165–80.
- Hornsby, J. (1997). "Truth: The Identity Theory," *Proceedings of the Aristotelian Society* 97: 1–24.
- Horwich, P. (1998). *Truth*, 2nd edn, revised. Oxford: Clarendon Press.
- Joachim, H. (1939) [1906]. *The Nature of Truth*. London: Oxford University Press.
- Kalderon, M. (1999). "The Transparency of Truth," *Mind* 106: 475–97.
- King, J. (2007). *The Nature and Structure of Content*. Oxford: Oxford University Press.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Lewis, D. (1986). *On The Plurality of Worlds*. Oxford: Blackwell.
- Lewis, D. (2001). "Forget about the 'Correspondence Theory of Truth,'" *Analysis* 61: 275–80.

- Lynch, M. (ed.) (2001). *The Nature of Truth*. Cambridge, MA: MIT Press.
- Mackie, J. (1973). *Truth, Probability and Paradox*. Oxford: Clarendon Press.
- McDowell, J. (1994a). "The Content of Perceptual Experience," *Philosophical Quarterly* 44: 190–205.
- McDowell, J. (1994b). *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, J. (2005). "The True Modesty of an Identity Conception of Truth: A Note in Response to Pascal Engel (2001)," *International Journal of Philosophical Studies* 13: 83–8.
- McGrath, M. (2003). "What the Deflationist May Say about Truthmaking," *Philosophy and Phenomenological Research* 66: 666–88.
- Meinong, A. (1983) [1902]. *On Assumptions*. Berkeley: University of California Press.
- Moffett, M. (2003). "Knowing Facts and Believing Propositions: A Solution to the Problem of Doxastic Shift," *Philosophical Studies* 115: 81–97.
- Moore, G. E. (1899). "The Nature of Judgment," in Baldwin (ed.) (1993), 1–19.
- Moore, G. E. (1901). "Truth and Falsity," in Baldwin (ed.) (1993), 20–2.
- Moore, G. E. (1953). *Some Main Problems of Philosophy*. London: George Allen & Unwin.
- Parsons, T. (1993). "On Denoting Propositions and Facts," *Philosophical Perspectives* 7: 441–60.
- Priest, Graham. (2000). "Could Everything be True?," *Australasian Journal of Philosophy* 78: 189–95.
- Prior, A. N. (1971). *Objects of Thought*. Oxford: Clarendon Press.
- Russell, B. (1937) [1903]. *The Principles of Mathematics*. London: George Allen & Unwin.
- Salmon, N. (1986). *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Salmon, N. (1989a). "Illogical Belief," *Philosophical Perspectives* 3: 243–85.
- Salmon, N. (1989b). "Tense and Singular Propositions," in J. Almog, H. Wettstein, and J. Perry (eds.), *Themes From Kaplan*. Oxford: Oxford University Press, 331–92.
- Shorter, J. M. (1962). "Facts, Logical Atomism and Reducibility," *Australasian Journal of Philosophy* 40: 283–302.
- Slote, M. (1974). *Metaphysics and Essence*. Oxford: Blackwell.
- Soames, S. (1985). "Lost Innocence," *Linguistics and Philosophy* 8: 59–71.
- Soames, S. (1987). "Direct Reference, Propositional Attitudes and Semantic Content," *Philosophical Topics* 15: 47–87.
- Soames, S. (1989). "Semantics and Semantic Competence," *Philosophical Perspectives* 3: 575–96.
- Sosa, E. (1993). "Epistemology, Realism, and Truth: The First Philosophical Perspectives Lecture," *Philosophical Perspectives* 7: 1–16.
- Stalnaker, R. (1976). "Propositions," in A. F. MacKay and D. D. Merrill (eds.), *Issues in the Philosophy of Language*. New Haven: Yale University Press, 79–91.
- Stoljar, D. and N. Damjanovic (2007). "The Deflationary Theory of Truth," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), <http://plato.stanford.edu/archives/spr2009/entries/truth-deflationary/>
- Strawson, P. F. (1950). "Truth," *Proceedings of the Aristotelian Society Supplementary Volumes* 24: 129–56 and reprinted in S. Blackburn and K. Simmons (eds.) (1999), 162–82. Page references are to the latter.
- Strawson, P. F. (1998). "Reply to John R. Searle," in L. Hahn (ed.), *The Philosophy of P. F. Strawson*. Chicago: Open Court.
- Vendler, Z. (1975). "On What We Know," in K. Gunderson (ed.) *Language, Mind and Knowledge*, Minnesota Studies in the Philosophy of Science 7: 370–90.

- Wetzel, T. (2003). "States of Affairs," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), <http://plato.stanford.edu/archives/fall2008/entries/states-of-affairs/>
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus*, first English edn (1922). trans. C. K. Ogden. London: Routledge and Kegan Paul.
- Woozley, A. D. (1949). *Theory of Knowledge*. London: Hutchinson.

CHAPTER 11

THE PRAGMATIST THEORY OF TRUTH

CHERYL MISAK

11.1 INTRODUCTION

PRAGMATISM originated in Cambridge, Massachusetts in the 1860s in The Metaphysical Club—a short-lived reading group whose members included William James and Charles Sanders Peirce. John Dewey, who was a graduate student of Peirce's at John Hopkins two decades later, became the third pillar of what is recognized as classical pragmatism.

At the heart of pragmatism is an account of what it is to properly understand a concept. The pragmatic maxim requires our concepts to be linked to experience and practice. All pragmatists, that is, want their philosophical explanations down-to-earth—natural as opposed to supernatural. As Peirce put it, we must look to the upshot of our concepts in order to understand them (*CP* 5.4). Pragmatism is committed to “taking a look,” as Ian Hacking has said—to keeping philosophy connected to first-order inquiry, to real examples, to action, and to real-life expertise (2007: 36).

Pragmatist accounts of truth must be understood as arising from this motivation. All pragmatists want to extract the concept of truth from our practices of inquiry, reason-giving, and deliberation. They also all want to be holists—they take their view of truth to encompass the whole of science, art, religion, ethics, and politics. No realm of inquiry is fenced off from the outset, although it might well be that some kinds of inquiry fail to meet the bar for truth-aptness.

It should be clear that pragmatism, of any stripe, will be set against versions of the correspondence theory of truth, on which a statement is true if it gets right or mirrors the human-independent world. For that concept of truth introduces an element that is unknowable by human inquirers and believers. That may be where the similarities between the varieties of pragmatist theories of truth end. There is an ongoing debate within pragmatism about the nature of its heart and soul. As Robert Westbrook says,

pragmatism has always been less a coherent philosophical school than a contentious family of thinkers holding distinct but related positions on the “workmanlike” nature of knowledge, meaning, and truth (2005: 1).

Pragmatism was an intensely debated view in the early 1900s. James’s *Pragmatism: A New Name for some Old Ways of Thinking* appeared in 1907, “cometlike on our intellectual horizon” (Carus 2001a [1911]: 44). The view shone brightly right through to John Dewey’s death in 1959, but then it seemed to burn out. In the 1970s, Richard Rorty brought it back under a spotlight, with an account based on the thought of James and Dewey.

Philosophy, James said, “is at once the most sublime and the most trivial of human pursuits. It works in the minutest crannies and it opens out the widest vistas” (1949 [1907]: 6). He was more a vista than a crannies man. Peirce was the most careful and analytic of the early pragmatists, perfectly happy working in the crannies as well as opening out the vistas. But he never found stable employment in academia and was hardly known as he toiled in poverty on his philosophy. His account of truth survived in the hands of C. I. Lewis and, surprisingly, Frank Ramsey. Lewis will not get the attention he deserves in this chapter, due to space constraints. But I will speak briefly to Ramsey’s advance on Peirce. For a more sustained discussion, see Misak (2016).

11.2 PEIRCE: TRUTH AND THE END OF INQUIRY

The pragmatic maxim, in Peirce’s hands, is designed to capture just one, albeit very important, aspect of what it is to understand something. Not only does one have to know how to give an analytic definition of a concept and how to pick out instances of it, but one has to know what to expect of beliefs containing the concept. If a belief has no consequences then it lacks a dimension we would have had to get right were we to fully understand it. The maxim cashes out in a prescription for philosophy: “we must not begin by talking of pure ideas,—vagabond thoughts that tramp the public roads without any human habitation,—but must begin with men and their conversation” (*CP* 8.112, 1900). We must connect our philosophical theories to that with which we have “dealings” (*CP* 5.416).

When Peirce turns this maxim on the concept of truth, the upshot is an aversion to “transcendental” accounts, such as the correspondence theory, on which a true belief is one that gets right or mirrors the believer-independent world. (*CP* 5.572). Such accounts of truth are examples of those “vagabond thoughts.” They make truth “the subject of metaphysics exclusively.” For the very idea of the believer-independent world, and the items within it to which beliefs or sentences might correspond, seem graspable only if we could somehow step outside of our corpus of belief, our practices, or that with which we have dealings.

Peirce thinks that the correspondence concept of truth is missing a connection with our practices. But he is perfectly happy with it as a “nominal” definition, useful to those who have never encountered the word before (*CP* 8.100). If we want a more robust or a full account of truth, we need to provide a pragmatic elucidation—an account of the role the concept plays in practical endeavors. David Wiggins nicely recaps the insight at the heart of Peirce’s pragmatism as follows. When a concept is “already fundamental to human thought and long since possessed of an autonomous interest,” it is pointless to try to define it. Rather, we ought to attempt to get leverage on the concept, or a fix on it, by exploring its connections with practice (2002: 316).

The concepts of belief and truth are examples of those concepts, fundamental to human thought, in which we have a long-standing autonomous interest. The application of the pragmatic maxim to belief leads Peirce to adopt a dispositional account of belief on which beliefs are in part “that upon which a man is prepared to act”; or “habits of mind,” which are “good or otherwise,” or “safe” or otherwise (*CP* 5.12; *W* 3.245). And with respect to truth, Peirce argues that if we are to bring the concept down to earth from metaphysical flights of fancy, we must see how it engages with our practices of assertion, inquiry, reasons, evidence, and belief. For those are the “dealings” connected to truth. When we assert, believe, or inquire, we take ourselves to be aiming at truth. We want to know, for instance, what methods might get us true belief; whether it is worth our time and energy to inquire into certain kinds of questions; whether a discourse such as moral discourse aims at truth or whether it is a radically subjective matter, not at all suited for truth-value.

“The Fixation of Belief” is one of the few papers Peirce managed to publish during his lifetime. In it, he sets out an account of inquiry and truth in a very provocative way. Inquiry, he says, is the struggle to rid ourselves of doubt and achieve a state of belief. He tells us that “the sole aim of inquiry” is to settle belief and that a belief which would be permanently settled is a true belief. Much of the rest of the paper is taken up with addressing the objection which wants to leap off the page: what if a belief was settled by “the rack” or by a totalitarian regime? Would such beliefs be true? Peirce is setting out here the most pressing problem for anyone who thinks that the concept of truth must be linked to our practices of inquiry. It is very hard to see how we can get a normative concept from a mere description of our practices. The following, in a nutshell, is Peirce’s solution to the problem.

Our body of background beliefs is susceptible to doubt on a piecemeal basis, if that doubt is prompted by surprising or recalcitrant experience—by “some positive reason” (*CP* 5.51). Cartesian doubts are “paper” or “tin” doubts—they cannot motivate inquiry. He says,

... there is but one state of mind from which you can “set out,” namely, the very state of mind in which you actually find yourself at the time you do “set out” a state in which you are laden with an immense mass of cognition already formed, of which you cannot divest yourself if you would ... Do you call it doubting to write down on a piece of paper that you doubt? If so, doubt has nothing to do with any serious business ... (*CP* 5.416)

The most obvious kind of positive reason for igniting doubt is a surprising experience—an experience that runs counter to what our settled beliefs lead us to expect. That is, we must regard our background beliefs as true, until experience throws one or some group of them into doubt. The inquirer “is under a compulsion to believe just what he does believe . . . as time goes on, the man’s belief usually changes in a manner which he cannot resist . . . this force which changes a man’s belief in spite of any effort of his may be, in all cases, called a *gain of experience*” (MS 1342: 2; emphasis in the original).

Peirce links this idea to the scientific method. It is the method which pays close attention to the fact that beliefs fall to the surprise of recalcitrant experience. Inquiry “is not standing upon the bedrock of fact. It is walking upon a bog, and can only say, this ground seems to hold for the present. Here I will stay till it begins to give way” (CP 5.589). Only science or the method that pays attention to the force of experience will provide us with stable ground on which to proceed. The specious methods (torture, totalitarian regimes, etc.) won’t fix belief that will really be settled—they will eventually be assailed by doubt. Similarly, the a priori method is a “failure,” because it

makes of inquiry something similar to the development of taste; but taste, unfortunately, is always more or less a matter of fashion . . . [And] I cannot help seeing that . . . sentiments in their development will be very greatly determined by accidental causes. Now, there are some people, among whom I must suppose that my reader is to be found, who, *when they see that any belief of theirs is determined by any circumstance extraneous to the facts, will from that moment not merely admit in words that that belief is doubtful, but will experience a real doubt of it, so that it ceases to be a belief.* (W 3.253; emphasis added)

Peirce’s argument is that a belief that would be permanently settled is indeed true. But it is very hard to really settle beliefs. For beliefs are such that they resign in the face of recalcitrant experience or in the knowledge that they were put in place by a method that did not take experience seriously. They resign in the fact of knowledge that they were determined by something accidental or extraneous to the facts. Another way of putting the point is that it is a constitutive norm of belief that a belief is responsive to the evidence and argument for or against it.

Hence, for Peirce, the inquirer is not merely after any old settled belief. He is after beliefs that are settled in a way that is connected with reasons and evidence for and against them. He is after beliefs that will serve him well in the future—beliefs that will not disappoint; that will guide action on a safe course; that will continue to fit with the experience, evidence, and argument. It is not so easy to end the irritation of doubt. It is not so easy to really fix belief. As Wiggins (2004) puts it, hard on the heels of the thought that truth is internally related to assertion or belief comes the thought that truth is also internally related to inquiry, reasons, evidence, and standards of good belief. If we unpack the commitments we incur when we assert and believe, we find that we have imported all these notions.

Peirce argues that when we think of how truth engages with our practices, we shall see that we need to think of a true belief as the very best that inquiry could do—a belief that would forever settle doubts; would be “indefeasible”; would not be improved upon; would never lead to disappointment; would forever meet the challenges of reasons, argument, and evidence. A true belief is a belief we would come to, were we to inquire as far as we could on a matter. He sometimes put this idea in the following unhelpful way: a true belief would be agreed upon at the hypothetical or “fated” end of inquiry (see *W* 3.273). But his considered and much better formulations are the ones above. A true belief would withstand doubt, were we to inquire as far as we fruitfully could into the matter. On the whole, he tries to stay away from unhelpful ideas such as the final end of inquiry, perfect evidence, and the like.

This is not to say that truth has now been *identified* as that which satisfies our aims in assertion and inquiry. We must be careful to not take these elucidations of truth to be attempts at analytic definition. Nothing could be clearer than Peirce’s intention to avoid that. A dispute about definition, he says, is usually a “profitless discussion” (*CP* 8.100). Again, Wiggins sees the point clearly: “To elucidate truth in its relations with the notion of inquiry, for instance, as the pragmatist does, need not . . . represent any concession at all to the idea that truth is *itself* an ‘epistemic notion’” (2002: 318; emphasis in the original). Once we see that the concepts of assertion, inquiry, and truth live in the same conceptual neighborhood, we can get a grip on the concept of truth by exploring the connections between it and its neighbors. This will not be an analysis of the essence of truth, but a way of getting clearer about what truth is.

One way of describing this project is to say that Peirce deflates the idea of truth by linking it to belief, assertion, experience, and inquiry. What we do when we offer a justification of “*p* is true” is to offer a justification for the claim that *p*. There is an unseverable connection between making an assertion and claiming that it is true. If we want to know whether it is true that Toronto is north of Buffalo, there is nothing additional to check on (“a fact,” “a state of affairs”)—nothing over and above our consulting maps, driving or walking north from Buffalo to see whether we get to Toronto, and so on. The question of the *truth* of the statement does not involve anything more than investigating the matter in our usual ways. For a claim’s fitting and continuing to fit with all the evidence and argument is all we can be interested in. Our attention must be on first-order inquiry into the claim itself, not on “philosophical” inquiry into the nature of truth. For the best kind of philosophical inquiry into the nature of truth draws out the connection between truth and the satisfaction of our aims in first-order assertion and inquiry.

The most significant challenge to Peirce’s account of truth is as follows: if a true belief is one that would be permanently resistant to doubt, what about those beliefs that are settled because no one cares to question them or because they are no longer accessible? Peirce calls these “buried secrets” and he spent considerable time trying to meet the challenge.

His response is that it is a regulative assumption of inquiry that, for any matter into which we are inquiring, we would find an answer to the question that is pressing on us. Otherwise, it would be pointless to inquire into the issue: “the only assumption upon

which [we] can act rationally is the hope of success" (W2.272). Thus the principle of bivalence—for any p , p is either true or false—rather than being a law of logic, is a regulative assumption of inquiry. It is something that we have to assume if we are to inquire into a matter. It is taken by logicians to be a law of logic by a "saltus"—by an unjustified leap. Peirce distinguishes his approach from that of the transcendentalist:

when we discuss a vexed question, we *hope* that there is some ascertainable truth about it, and that the discussion is not to go on forever and to no purpose. A transcendentalist would claim that it is an indispensable "presupposition" that there is an ascertainable true answer to every intelligible question. I used to talk like that, myself; for when I was a babe in philosophy my bottle was filled from the udders of Kant. But by this time I have come to want something more substantial. (CP 2.113)

Not only does the fact that an assumption is indispensable to our practice of inquiry not entail that it is a necessary truth; it does not even entail that it is true or even that it ought to be believed. Peirce says, "I do not admit that indispensability is any ground of belief. It may be indispensable that I should have \$500 in the bank—because I have given checks to that amount. But I have never found that the indispensability directly affected my balance, in the least" (CP 2.113). We must make these assumptions "for the same reason that a general who has to capture a position or see his country ruined, must go on the hypothesis that there is some way in which he can and shall capture it" (CP 7.219).

Thus Peirce's argument is that if we are to inquire rationally about some particular issue, then we must assume that there is at least a chance of there being an upshot to our inquiry. We must also assume that there is a reality independent of our beliefs about it, whose character we can discover. We must also assume that there is an explanation for what we observe. Refusing to make these assumptions is to block the path of inquiry and, in Peirce's books, that is the cardinal philosophical sin. Our reason for making the assumptions is driven, Peirce says, by "desperation." If we do not make them, we will "be quite unable to know anything of positive fact" (CP 5.603). Faced with an assumption without which we cannot continue in a practice of utmost importance, we must embrace it, "however destitute of evidentiary support it may be" (CP 7.219).

Peirce, like all the early pragmatists, is a holist in that he thinks that every area of inquiry falls under the scope of truth and knowledge. (He is not the kind of holist, though, who thinks that a whole theory is the bearer of truth.) His treatment of mathematical and logical beliefs provides a good illustration of what he was trying to achieve. The history of empiricism is littered with attempts to show how these statements, although not obviously connected to experience, are nonetheless legitimate. Most of those attempts (think of Hume and the Vienna Circle) trade on the idea that mathematics and logic are somehow exempt from the rigors of the empiricist criterion. But Peirce treats mathematics and logic as a whole with the rest of genuine inquiry.

He argued that mathematical and logical hypotheses *are* connected to experience in the requisite way. They meet the requirement set out in the pragmatic maxim: we expect certain things to be the case if they are true. Not only might we have practical or applied

or bridge-building expectations about mathematics, but even hypotheses in pure mathematics have consequences. They have consequences, Peirce argued, in diagrammatic contexts. Diagrams provide us with a forum for matters to impinge upon us.

Peirce put considerable effort into trying to get this thought right. (Indeed, he developed a first-order quantified logic with a diagrammatic proof system just as Frege was developing his logic.) In 1905, he suggested that there are two kinds of experience: ideal and real. The latter is sensory experience and the former is experience in which “operations on diagrams, whether external or imaginary, take the place of the experiments upon real things that one performs in chemical and physical research” (CP 4.530). But already in 1872, this idea had a central place in his thought. Mathematical and logical inquiry

involves an element of observation; namely, [it] consists in constructing an icon or diagram the relation of whose parts shall present a complete analogy with those of the parts of the object of reasoning, of experimenting upon this image in the imagination, and of observing the result so as to discover unnoticed and hidden relations among the parts. (W3.41)

The mathematician’s “hypotheses are creatures of his own imagination; but he discovers in them relations which surprise him sometimes” (CP 5.567). This surprise is the force of experience.

Sometimes Peirce distinguishes the two kinds of experience by saying that everyone inhabits two worlds: the inner (or the ideal) and the outer (or the real). We react with the outer world through a clash between it and our senses. We react with the inner world by performing thought experiments. Inquiry, he says, has

two branches; one is inquiry into Outward Fact by experimentation and observation, and is called *Inductive Investigation*; the other is inquiry into Inner Truth by inward experimentation and observation and is called *Mathematical* or *Deductive Reasoning*. (MS 408: 150)

The distinction between these two kinds of experience and two kinds of inquiry is not, however, hard and fast. External facts are simply those which are “ordinarily regarded as external while others are regarded as internal” (W2.205). The inner world exerts a comparatively slight compulsion upon us, whereas the outer world is full of irresistible compulsions. Nonetheless, the inner world can also be “unreasonably compulsory” and have “its surprises for us” (CP 7.438).

It may seem that Peirce’s assumption that there is a reality independent of beliefs invites the claim that he really holds some kind of correspondence theory. But this would be a mistake, for the assumption is one made *within* inquiry. It is not being offered as the beginnings of a theory of truth per se.

One thing that needs attention is the matter of saying just which kinds of inner experiences count and which do not. Peirce struggled with this problem and it would

be silly to say that he got it right. But we shall see that one thing is clear. He thought that James got it wrong.

11.3 WILLIAM JAMES: TRUTH AND SATISFACTION

William James, like the logical empiricists after him, has his pragmatic maxim making short work of many long-standing and seemingly intractable metaphysical problems: “If no practical difference whatsoever can be traced, then the alternatives mean practically the same thing, and all dispute is idle” (1949 [1907]: 45). But he parts company from the logical empiricists in a dramatic way when he tells us that the kind of difference he is talking about is any kind of difference to you or to me:

There can be no difference anywhere that doesn’t make a difference elsewhere—no difference in abstract truth that doesn’t express itself in a difference in concrete fact and in conduct consequent upon that fact, imposed on somebody, somehow, somewhere, and sometime. The whole function of philosophy ought to be to find out what definite difference it will make to you and me, at definite instants of our life, if this world-formula or that world-formula be the true one. (1949 [1907]: 30)

Whereas the logical empiricists have all hypotheses requiring empirical consequences, James has all hypotheses requiring consequences that include those for my conduct or for your conduct. One wonders whether any hypothesis will fail to make the grade.

When James turns his maxim on the concept of truth, it is unsurprising that the result is rather radical. He sets out his view on truth and objectivity thus: “Any idea upon which we can ride . . . any idea that will carry us prosperously from any one part of our experience to any other part, linking things satisfactorily, working securely, simplifying, saving labor, is . . . true *instrumentally*” (1949 [1907]: 34). “Satisfactorily,” for James, “means more satisfactorily to ourselves, and individuals will emphasize their points of satisfaction differently. To a certain degree, therefore, everything here is plastic” (1949 [1907]: 35). The individuality or subjectivity built into James’s version of the pragmatic maxim is clearly manifested in his account of truth.

James rather infamously argued in “The Will to Believe” that if the belief in God made a positive or a happy impact on someone’s life, then it reasonably could be taken as true by that person. Religious hypotheses, like all hypotheses, need to be verified. But verification can involve finding out whether the hypothesis has good effects on the believer’s life. Similarly, the matter of the truth of the hypothesis involves finding out whether the hypothesis works nicely for the believer:

If religious hypotheses about the universe be in order at all, then the active faiths of individuals in them, freely expressing themselves in life, are the experimental tests by

which they are verified, and the only means by which their truth or falsehood can be wrought out. The truest scientific hypothesis is that which, as we say, "works best"; and it can be no otherwise with religious hypotheses. (1979 [1897]: 8)

The objection to this line of thought is that such evidence may be relevant to the question of whether or not religion is good for human beings, but not relevant to the question of whether God exists. This is the very objection that Peirce put to James. The book *The Will to Believe* is dedicated "To My Old Friend, Charles Sanders Peirce, To whose philosophic comradeship in old times and to whose writings in more recent years I owe more incitement and help than I can express or repay." Peirce was touched, but nonetheless, he doesn't have much good to say about James's essay. He tells James in a 1909 letter: "I thought your *Will to Believe* was a very exaggerated utterance, such as injures a serious man very much . . ." (CL 12: 171) He thinks that James's view amounts to: "Oh, I could not believe so-and-so, because I should be wretched if I did" (CP 5.377).

James's account of truth has inspired much vitriol and can be seen as responsible for much of pragmatism's bad reputation. But a few words of caution are required. First, a careful reading of James shows first that when he argues against the idea that truth is static and asserts, rather, that it is plastic and that "truth *happens* to an idea" (1949 [1907]: 97), what he is quite clearly talking about is not truth, but what we *take* to be true. In *Pragmatism* he says that "the great assumption of the intellectualists is that truth means essentially an inert static relation. When you've got your true idea of anything, there's an end of the matter. You're in possession; you *know*; you have fulfilled your thinking destiny" (1949 [1907]: 96). James, with Peirce, wants to correct this mistaken assumption. When you have a settled, well-grounded belief, you don't *know*; you don't have an end to the matter. For inquiry might well overturn your belief. That is, sometimes James's thoughts about the plasticity of truth are merely sensible statements of fallibilism about belief.

Second, James sometimes makes it clear that he was concerned to characterize truth as something that was of human value, without making a true belief whatever this or that human finds valuable at this or that time. He tries to correct a misunderstanding of his position by arguing that, contrary to his critics, he holds that the true is "the expedient," but the expedient "in the long run and on the whole, of course" (1909 [1975]: 4). That is, James, with Peirce, wants to argue that true beliefs are beliefs which survive because they *deserve* to survive, not because they happen to survive for this or that person. We start with our background beliefs and revise in light of good reasons and the force of experience.

In *Pragmatism*, James offers us a clutch of metaphors for the growth of knowledge. One of them likens the change in belief to house renovations: "You may alter your house *ad libitum*, but the ground plan of the first architect persists—you can make great changes, but you cannot change a Gothic church into a Doric temple" (1907 [1975]: 83). Another ties the growth of knowledge to a different kind of contingent residue: "You may rinse and rinse the bottle, but you can't get the taste of the medicine or whisky that first filled it wholly out" (1907 [1975]: 83). These metaphors employ the pragmatist theme

that “we patch and tinker” (1907 [1975]: 83). They cohere with Peirce’s metaphor that the inquirer is standing on a bog and only moves forward when the ground underneath him begins to give way. We start from where we find ourselves in inquiry and move forward from there, laden with beliefs and frameworks that were put in place by previous generations of inquirers.

But fair or unfair, the vitriol about James set the tone for how pragmatism was viewed for decades to come. James’s critics latched on to the most simple and clear statements of his view—that we make truth and that a true belief is one that is useful or works. Bertrand Russell and G. E. Moore led the charge.

Russell turned James’s account of truth on itself, noting that if James’s theory of truth is itself to be useful, there must be a way of telling when the consequences of a belief are useful or good (1992 [1966]: 201):

We must suppose that this means that the consequences of entertaining the belief are better than those of rejecting it. In order to know this, we must know what are the consequences of entertaining it, and what are the consequences of rejecting it; we must know also what consequences are good, what bad, what consequences are better, and what worse.

This, of course, is a very tall order, which Russell immediately illustrates with two examples. First, the consequences of believing the doctrine of the Catholic faith might make one happy “at the expense of a certain amount of stupidity and priestly domination” (1992 [1966]: 201). It is unclear how we are to weigh these benefits (such as happiness) and burdens (such as priestly domination) against each other. Second, the effects of Rousseau’s doctrines were far-reaching—Europe is a different place from what it would have been without them. But how do we disentangle what the effects have been? And even if we could do that, whether we take them to be good or bad depends on our political views. The question of whether the consequences of believing something are on the whole good or bad is an extraordinarily difficult one.

In a related objection, Russell points that one can take “works” or “pays” in two very different ways. In science, a hypothesis works if we can deduce a number of verifiable hypotheses from it. But for James, a hypothesis works if “the effects of believing it are good, including among the effects . . . the emotions entailed by it or its perceived consequences, and the actions to which we are prompted by it or its perceived consequences.” This is a totally different conception of “working,” and one for which the authority of scientific procedure cannot be invoked (1992 [1966]: 210).

Moore angrily reviewed James’s *Pragmatism* in the 1907 *Proceedings of the Aristotelian Society*. Here is a catalogue of his objections. First, he points to a problem that dogs all pragmatist views of truth. If truth is tightly connected to what we can verify, what do we think of statements for which the evidence has been destroyed; or statements that are so trivial that no one has bothered to collect any evidence for them; or statements the evidence for which lies buried deep in the past? (1992 [1907]: 165, 179). We have seen that

Peirce offered a solution to this problem, unbeknownst to the rest of the philosophical world. James does not tackle it.

Second, with Russell, Moore interrogates the linkage between the true and the useful. If usefulness is a property that may come and go (in James's own words), then Moore says that "a belief, which occurs at several different times, may be true at some of the times at which it occurs, and yet untrue at others" (1992 [1907]: 183). The truth of a belief, that is, seems to vary from time to time and from culture to culture. Truth is not a stable property of beliefs. That is an anathema, in Moore's view.

Third, Moore takes on James's claim that we make the truth: "I think that he certainly means to suggest that we not only make our true beliefs, but also that we *make* them true" (1992 [1907]: 191; emphasis in the original). Moore thinks that it is crazy to suggest that my belief that *p* makes it true that *p*. My (correct) belief that it rained today did not make it rain today.

One can see that under a barrage of well-formed criticism such as this, pragmatism's reputation across the Atlantic was bound to suffer. But it came under similar stress in America. Paul Carus in his 1911 *Truth on Trial* argues that Jamesian pragmatism replaces "the belief in the stability of truth, in its persistence and eternality" with "a more elastic kind of truth which can change with the fashions and makes it possible that we need no longer trouble about inconsistencies; for what is true to one need no longer be true to others, and the truth of to-day may be the real now, and yet it may become the error of the to-morrow" (2001a [1911]: 110). Hence, he thinks that pragmatism has put truth on trial, with James as the hapless prosecutor.

11.4 JOHN DEWEY: TRUTH AND THE RESOLUTION OF PROBLEMATIC SITUATIONS

John Dewey, in good pragmatist form, also links truth with inquiry. His 1938 *Logic: The Theory of Inquiry* is the culmination of decades of thinking about the relationship. He starts off by accepting Peirce's doubt-belief model of inquiry, duly acknowledged:

Doubt is uneasy; it is tension that finds expression and outlet in the processes of inquiry. Inquiry terminates in reaching that which is settled. This settled condition is a demarcating characteristic of genuine belief. In so far, belief is an appropriate name for the end of inquiry. (LW 12: 15)

Inquiry is the resolving of a problematic situation. No one investigating a particular problem aims for certainty or absolute truth. Rather, they aim at belief, or security, or a reliable solution to the problem at hand.

With Peirce and James, Dewey argues that we are always immersed in a context of inquiry, where the decision to be made is a decision about what to believe from here, not what to believe were we able to start from scratch—from certain infallible foundations. Although Dewey sometimes says that “truth is the end and standard of thinking” (MW 4: 64), on the whole he prefers “warranted assertibility” as that at which we aim or that which closes off a particular inquiry (LW 12: 15).

Dewey tried to bring the external world into his pragmatism, but an awkward metaphysics is the result. Experience, he thinks, is “an affair of the intercourse of a living being with its physical and social environment” (MW 10: 6). He says, “When we experience something we act upon it, we do something with it; then we suffer and undergo the consequences” (MW 9: 146). The problematic situation that calls for inquiry is a situation in which the human organism interacts with “nature” (LW 12: 110). It is the “interaction of organic responses and environing conditions” (LW 12: 111).

As Robert Talisse (2002) has critically noted, for Dewey, it is the situation which has the doubt, not the inquirer. Dewey himself could not be clearer: “It is the *situation* that has these traits. We are doubtful because the situation is inherently doubtful” (LW 12: 109; emphasis in the original) Ernst Nagel expressed a worry about this metaphysics in 1929:

The question naturally arises how Professor Dewey comes to have a metaphysics. How does he *know* that specificity, interaction, change, characterize all existence, and that these distinctions are not merely logical, made for purposes of getting along in this world, but characters of an independent existence? Why does he impute the features presented in human experience to a nature embracing, but containing more than, that experience? (1929: 455)

Even James says that Dewey and his disciples have “a peculiar view of “fact” on which

A fact and a theory have not different natures, as is usually supposed, the one being objective, the other subjective. They are both made of the same material, experience-material, namely, and their difference relates to their way of functioning solely . . . It is “fact” when it functions steadily; it is “theory” when we hesitate. (1977 [1904]: 104))

James, of course, will be more sympathetic than most. But even James sees that one of the “great gaps” in Dewey’s system is that there is “no account of the fact . . . that different subjects share a common object-world.” (1977 [1904]: 105)

Nonetheless, one can see Dewey struggling to find a position that combines the best of Peirce and James. He distinguishes his view from James’s by making it clear that he does not go in for the problematic idea that “any good which flows from the acceptance of a belief” can be treated as “evidence” for the truth of that belief (MW 4: 109). He says,

Since Mr. James has referred to me as saying “truth is what gives satisfaction,” I may remark . . . that I have never identified any satisfaction with the truth of an idea,

save *that* satisfaction which arises when the idea as working hypothesis or tentative method is applied to prior existences in such a way as to fulfill what it intends. (MW 4: 109; emphasis in the original)

Dewey is alert to how James moves too easily between “since true ideas are good, any idea if good in any way is true” (MW 4: 108). But Dewey is a careful reader of James and takes “his real doctrine” to be something more like what Dewey himself is trying to establish: “a belief is true when it satisfies both personal needs and requirements of objective things” (MW 4: 112). Personal needs or human psychology, Dewey holds, cannot be kept out of questions about truth. Better to recognize it, control it, and accept responsibility for it, than to try to ignore it (MW 4: 114).

With Peirce and James, Dewey is set against the “intellectualist” account of truth, on which truth is “*antecedent* to any process of verification” (MW 4: 76). The problem with the intellectualist is that he cannot make sense of truth as the internal property of an idea hooking onto an idea-independent reality. The “dualism” between mind and matter or consciousness and objects “where each of these terms is supposed to refer to some fixed order of existence” is an unbridgeable gap (MW 4: 79). He says,

Then, of course, comes up the question of the nature of the agreement, and of the recognition of it. What is the experience in which the survey of both idea and existence and their agreement recognized? Is it an idea? . . . Then what has become of the postulate that truth is agreement of idea with existence beyond idea? Is it an absolute which transcends and absorbs the difference? Then . . . what is the test of any specific judgment? What has become of the correspondence of fact and thought? Or, more urgently, since the pressing problem of life, of practice and of science is the discrimination of the relative, or superior, validity of this or that theory, plan or interpretation, what is the criterion of truth . . . ? (MW 4: 79)

Truth cannot be correspondence of a thought with how things “really are” (MW 6: 34). For we cannot get at how things really are: “we seem to require a third medium in which the original proposition and its object are surveyed together, are compared and their agreement or disagreement seen” (MW 6: 34).

None of the usual theories of truth stands up. No wonder, Dewey says, that “a third party has finally been rash enough to intervene” (MW 6: 33). That third party is the pragmatist, who does not look for a “ready-made” truth, but rather, looks toward the future (MW 6: 38).

Dewey was one of the classical pragmatists who tried to work out how ethics might come under the scope of truth and knowledge (Lewis being the second). Everything is brought under the umbrella of inquiry. A trial, for instance, is a problematic situation requiring inquiry (LW 12: 123). Its formal conceptions (misdemeanor, crime, torts, contracts, etc.) arise out of “ordinary transactions”—they are not imposed “from on high or from any external or *a priori* source” (LW 12: 106). “But when they are formed, they are also *formative*; they regulate the proper conduct of the activities out of which they develop” (LW 12: 106; emphasis in the original). When problems arise, the formal

theory that has been developed is brought to the matter and the theory itself may be revised in light of new insights prompted by the problem.

Dewey is offering us a unified theory of inquiry—a way of thinking about how we resolve problematic situations in science, ethics, politics, law, and art. The pragmatist, he says, “has at least tried to face, and not to dodge, the question of how it is that moral and scientific ‘knowledge’ can both hold of one and the same world” (MW 4: 132). “The problems of men” must be brought under the sweep of science or inquiry:

if we can discover ethical principles, these ought to give some guidance for the unsolved problems of life which continually present themselves for decision. Whatever may be true for other sciences it would seem that ethics at least ought to have some practical value . . . Man must act; and he must act well or ill, rightly or wrongly. If he has reflected, has considered his conduct in the light of the general principles of human order and progress, he ought to be able to act more intelligently and freely, to achieve the satisfaction that always attends on scientific as compared with uncritical or rule-of-thumb practice. (MW 5: 10)

The way that moral inquiry works is via intuition or thought experimentation:

It is a process of tentative action: we “try on” one or another of the ends, imagining ourselves actually doing them, going, indeed, in this make-believe action just as far as we can without actually doing them. In fact, we often find ourselves carried over the line here; the hold which a given impulse gets upon us while we are “trying it on” passes into over act without us having consciously intended it . . . Decision, resolution, the definitely formed plan, is the proper outcome of consideration. (EW 4: 251)

Our lives are full of problematic situations, some of them moral. We propose potential solutions to the problems which press on us, try to predict the consequences of the solutions’ implementations, and ask whether our reactions to those consequences would be positive or negative. We then test the solution that has withstood the challenge of thought experiment or experiment in the imagination. That is, we then see what the results actually are.

It is unsurprising that in ethics, too, Dewey has difficulties in saying how something external impinges upon us. A moral problem or situation presents itself to the individual as thus: “Which shall he decide for and why? The appeal is to himself; what does *he* really think the desirable end? What makes the supreme appeal to him? What sort of an agent, a person, shall he be? This is the question finally at stake in any genuinely moral situation: what shall the agent *be*?” (MW 5: 194).

The problem for this view is that it is not clear how it can be normative. Jennifer Welchman articulates the issue nicely:

One might object, however, that although ordinary deliberation can proceed along the lines Dewey suggests, the reflective deliberation that issues in remorse, regret, and the effort to reform oneself cannot. An act that is “right” because it is true to myself may still be wrong because the sort of self I am is a wrong or bad self. (1995: 109)

This is a specific statement of the general problem that dogs pragmatist theory of truth. How can the pragmatist avoid the naturalistic fallacy—the fallacy of trying to derive an “ought” from an “is” or trying to get normative conclusions from descriptive premises? Dewey sees the problem clearly: we are “face to face . . . with the only serious question a . . . wisely pragmatic philosophy need fear” (MW 6: 54). It is the problem of taking what is the case and making it normative: “equating . . . the intellectually satisfactory with the personally agreeable, or . . . the authentic with what happens to be authorized, . . . the legitimate with the legal” (MW 6: 5).

Dewey’s solution to the problem in *The Study of Ethics* is that we can tell that a conclusion is misguided even when we are aiming at self-realization, for the result will be internal conflict, discomfort, and “compunctions” (EW 4: 297). And we can query our desires, aims, and conceptions of ourselves once we see how they lead to disharmony. But this seems less than adequate. Not all evil-doers, for instance, appear to experience such discomfort. Later, he offers a different solution. When we try to solve problems we

convert strife into harmony, monotony into a variegated scene, and limitation into expansion. The converting is progress, the only progress conceivable or attainable by man. Hence every situation has its own measure and quality of progress, and the need for progress is recurrent, constant. (MW 14: 195)

In the 1920 *Reconstruction in Philosophy* he puts it thus:

the process of growth, of improvement and progress, rather than the static outcome and result, becomes the significant thing. Not health as an end fixed once and for all, but the needed improvement in health—a continual process—is the end and good. The end is no longer a terminus or limit to be reached. It is the active process of transforming the existent situation. Not perfection as a final goal, but the ever-enduring process of perfecting, maturing, refining is the aim in living . . . Growth itself is the only moral “end.” (MW 12: 181)

Meliorism is thus at the very heart of his position. Like Peirce, he thinks although certainty is not to be had, we go forward, in the hope that we do better and better. But on Peirce’s view, there is something that we are aiming at, despite the fact that we can never be certain that we have it: we are aiming at a belief which would forever stand up to experience and deliberation. On Dewey’s view, we are not aiming at even that endpoint. So it is not at all clear what “growth” or “progress” can mean. We struggle to find a robust enough conception of a right answer, or a better outcome, or a real solution to a problematic situation in his account of truth.

11.5 FRANK RAMSEY: TRUTH AND SUCCESS

Ramsey got his hands on Peirce’s work through C. K. Ogden, who had brought out Peirce’s first volume of posthumously collected papers alongside the 1923 American

edition. He was heavily influenced by it. In the manuscript he was working on when he died at the age of twenty-six in 1930, he is clear that the Jamesian position is not the whole or the best of pragmatism, saying, "What is ludicrous, is not the general idea" of pragmatism but "the way in which William James confused it especially in its application to religious beliefs" (OT: 91). Ramsey was in the first instance attracted to Peirce's dispositional account of belief. This core insight of pragmatism is the spur for his best-known result. In the 1926 "Truth and Probability," he showed that we can measure *partial* belief by seeing how people would act, especially in betting contexts. That is one way that Ramsey improves on Peirce's view.

Another way he improves on Peirce is as follows. In "General Propositions and Causality," Ramsey argues that generalization and conditionals cannot be analyzed, respectively, as infinite conjunctions and in terms of truth-values. Rather, they are habits or rules with which we "meet the future" (GC: 149). If I believe that all men are mortal, I adopt a rule or a habit of the form: if I meet a Φ , I shall regard it as a ψ (GC: 149). When I accept a conditional "if p then q ," I commit myself to acquiring the disposition to judge q whenever I judge p .

Ramsey starts to work through the tricky issue of how these habits can be "cognitive attitudes": "in what way can [such a habit] be right or wrong?" (GC: 146–7). The belief that all men are mortal will play out in diverse ways—I will be disposed to assert and affirm that all men are mortal in appropriate circumstances; I will drive my car carefully around those pedestrians I wish to remain alive; I will think that every person I meet will at some point die; I will not treat myself as an immortal exception; I may despair about the meaning of life; and so on. And my belief or rule or habit can be *evaluated* in terms of whether it manifests itself in appropriate ways (whether I adopt dispositions such as the ones above) and whether it continues to cohere with experience.

Similarly, a conditional with an unfulfilled antecedent can be evaluated. If a man has a cake and decides not to eat it because he thinks it will make him ill, we can judge him mistaken even if he does not eat the cake. We have different "degrees of expectation" as to the outcome, and we can "introduce any fact we know, whether he did or could know it" (GC: 155). Let's say he knew that I carefully baked the cake; that I'm an excellent baker; that I know he has no food allergies or aversions; and that I bear no ill will toward him. Then we might judge that he is irrational. If all these things hold, but he does not know them, then we might judge him mistaken.

We can also evaluate these attitudes because, as Ramsey puts it, they form *the system* with which we meet the future. If you and I meet the future with different systems, then we *disagree*. The future, we assume, will be compatible with one of our systems but not the others:

We do, however, believe that the system is uniquely determined and that long enough investigation will lead us all to it. This is Peirce's notion of truth as what everyone will believe in the end. (GC: 161)

We have habits of belief, and those beliefs are evaluated in terms of whether they serve us well, in a robust sense of “serve.” Pragmatism, for Peirce and Ramsey, is the position that beliefs should be evaluated based on both hindsight (whether the belief-formation method was connected to the facts) and foresight (whether the belief continues to work, fitting with future experience, other well-grounded beliefs, and enabling successful action). That, I submit, is the most compelling of the various pragmatist accounts of truth.

11.6 QUINE: HOLISM AND DEFLATIONISM

Willard Van Orman Quine arrived at Harvard in 1930 as a graduate student in philosophy, with a BA in mathematics. Two of his courses were taught by C. I. Lewis and it was here that he acquired his introduction to pragmatism. Quine was never terribly interested in the history of pragmatist ideas. He told Morton White, for instance, that reading Josiah Royce was like going through muck (White 1999: 121–4). Rather, Quine was one of those philosophers who took his pragmatist heritage and melded it into his free-standing philosophy, ushering in a new kind of analytic pragmatism.

Although Quine adopts as his theory of truth Tarskian deflationism (there is nothing more to “*p* is true” than “*p*”), he comes very close to putting forward a pragmatist account of truth. Indeed, Quine was initially happy to be placed in the pragmatist camp. In the abstract of his famous “Two Dogmas of Empiricism,” he asserts that one upshot of the paper is “a shift toward pragmatism.” He talks about science as being a tool for managing the flux of experience that could be right out of James and which could not have failed to bring Dewey to mind:

As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries—not by definition in terms of experience, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. For my part I do, qua lay physicist, believe in physical objects and not in Homer’s gods . . . But in point of epistemological footing the physical objects and the gods enter our conceptions only as cultural posits. The myth of physical objects is superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience. (1980 [1951]: 44)

But his relationship to pragmatism is complex and soon afterwards we Quine him distancing himself from the position.

Quine is responsible for entrenching the term “holism” into philosophical vocabulary. He argues that our entire belief system must be seen as one interconnected web.

Mathematics and logic are at the center, gradually shading into the theoretical sentences of science, and then to specific observation sentences at the periphery. When faced with recalcitrant experience, we must choose where to make adjustments in our web of belief. No kind of sentence in that web of belief is immune from revision.

Quine recoiled from extending his holism as far as Dewey was prepared to go—he was leery of the idea that ethics might form part of our web of belief: “apart from a salient marker or two” one finds “uncharted moral wastes” (1987: 5). But Quine’s early characterization of experience is certainly broad enough to bring moral judgments into the fold. Observation sentences

can be roughly distinguished from others by a behavioural criterion, involving no probing of sensations. For this is characteristic of them: witnesses will agree on the spot in applying an observation term, or in assenting to an observation sentence . . . (2008 [1975]: 230)

This leaves room for the possibility that some ethical statements will be observational—“That’s odious,” on seeing a sexual assault upon a child, for instance.

Quine, with his pragmatist predecessors, is a naturalist and anti-foundationalist. He made famous Neurath’s metaphor regarding the growth of knowledge. We are like sailors adrift at sea, never able to return to drydock to reconstruct our boat out of the finest materials. We work with what we have, replacing our boat of knowledge plank by plank, as required by the surprise of experience. The resonances with the metaphors offered by the early pragmatists are striking:

The naturalistic philosopher begins his reasoning within the inherited world theory as a going concern. He tentatively believes all of it, but believes also that some unidentified portions of it are wrong. He tries to improve, clarify and understand the system from within. (1981: 72)

That is a perfect summary of the pragmatism that Peirce was so keen to articulate. We go on our settled beliefs until some surprising experience throws them into doubt. Then we revise until we have another, better, settled belief upon which to rely.

11.7 RICHARD RORTY: TRUTH AND SOLIDARITY

Quine’s reluctance to embrace the pragmatist label left a space in the intellectual landscape—a space which Richard Rorty moved to fill in the 1970s. Rorty’s versions of the pragmatist theory of truth “tend to center” around those of James (Rorty 1995: 71).

Unsurprisingly, the objections that were leveled against James are the challenges faced by Rorty.

Rorty's rendering of the pragmatic maxim is that our concepts cannot outrun our practices or our current and ongoing "conversations" in which we must form our beliefs, make our decisions, and live our lives:

there is nothing deep down inside us except what we have put there ourselves, no criterion that we have not created in the course of creating a practice, no standard of rationality that is not an appeal to such a criterion, no rigorous argumentation that is not obedience to our own conventions. (1982: xlii)

Rorty argues that there is no practical difference between aiming to hold true beliefs and aiming to hold justified beliefs, and hence it makes no sense to speak of truth as a goal of inquiry. The inquirer can't compare her beliefs with reality—she can only try to live up to her epistemic responsibilities or the standards of her epistemic community. The yearning for an unconditional, impossible, indefinable, sublime thing like truth comes at the price of "irrelevance to practice" (2000: 2).

Rorty goes on to maintain that truth is simply solidarity; or what we have come to take as true; or what our peers will let us get away with saying (1979: 176). We must "substitute the idea of 'unforced agreement' for that of 'objectivity'" (1991: 38) in every domain of inquiry—in science as well as in morals and politics (1991: 36).

As with James, it looks as if the bar is being set very low. Any hypothesis that is agreed upon, any hypothesis we can get away with asserting, is as good as true. In 1909, James Pratt nailed the matter on the head with respect to Jamesian pragmatism. It

seeks to prove the truth of religion by its good and satisfactory consequences. Here, however, a distinction must be made; namely between the "good," harmonious, and logically confirmatory consequences of religious concepts as such, and the good and pleasant consequences which come from believing these concepts. It is one thing to say a belief is true because the logical consequences that flow from it fit in harmoniously with our otherwise grounded knowledge; and quite another to call it true because it is pleasant to believe. (1909: 186–7)

The difference between the view of Peirce, on the one hand, and the view of James and Rorty, on the other, can be nicely summarized by Pratt's distinction, with one caveat. Peirce holds that "a belief is true because the logical consequences that flow from it fit in harmoniously with our otherwise grounded knowledge" and James and Rorty hold (or at least fail to block the thought) that a belief can true "because it is pleasant to believe." The caveat is that Peirce insisted on a subjunctive formulation: a belief is true if the logical consequences *would* fit harmoniously with our otherwise grounded knowledge, were we to pursue our investigations as far as they could fruitfully go.

11.8 WAYS FORWARD

The debate within pragmatism continues to this day. It is a debate between those who take pragmatism to suggest that there is no truth and objectivity to be had anywhere and those who take pragmatism toward promise an account of truth that preserves our cognitive aspiration to getting things right. On the one side of the debate we have Rorty arguing, with his classical predecessor James, that there is no truth at which we might aim—only agreement within a community or what works best for an individual. On the other side of the divide, we have Ramsey and his classical predecessor Peirce, arguing that we should not look for a transcendental or metaphysical theory of truth, but we should nonetheless try to do justice to the objective dimension of human inquiry—to the fact that those engaged in deliberation and investigation take themselves to be aiming at successful action, getting things right, avoiding mistakes, and improving their beliefs and theories. Dewey and his successors, I suggest, try to fall in the middle—they would like to do justice to the objective dimension of inquiry, but they have not given themselves the wherewithal to do so.

On the more objective kind of pragmatism, the fact that our inquiries are historically situated does not entail that they lack objectivity. Neither does the fact that standards of objectivity themselves come into being and evolve. The trail of the human serpent is over everything (to use James's phrase), but as Peirce and Ramsey saw, this does not toss us into a sea of arbitrariness, where there is no truth or where truth varies from person to person and culture to culture.

REFERENCES

[*Note on references to the work of Peirce, Dewey, and Ramsey (full bibliographical details below):* *The Writings of Charles S. Peirce: A Chronological Edition* are cited as "W n.m; year," where n is the volume, m the page number. Peirce's *Collected Papers* are cited as "CP n.m," where n is the volume number, m the paragraph number. Peirce's manuscripts are cited as "MS n," where n is the manuscript number. All citations of John Dewey refer to his *Collected Works* (= J. A. Boydston, ed., 1967–91). *The Early Works of John Dewey, 1882–1898*, are cited as "EW n: m," where n is the volume number, and m the page number. *The Middle Works of John Dewey, 1899–1924*, are likewise cited as "MW n: m," and *The Later Works of John Dewey, 1925–1953* cited as "LW n: m." Ramsey's *On Truth* is cited as "OT," "General Propositions and Causality," in *Philosophical Papers* is cited as "GC."] "CL" refers to Skrupselis and Berkeley (eds.), *The Correspondence of William James*.

Boydston, J. A. (ed.) (1967–91). *The Collected Works of John Dewey, 1882–1953*. 37 vols. Carbondale, IL: Southern Illinois University Press.

James, W. (1949) [1907]. *Pragmatism: A New Name for Some Old Ways of Thinking*. New York: Longmans, Green.

James, W. (1975) [1909]. *The Meaning of Truth, in The Works of William James*, vol. 2. Cambridge, MA: Harvard University Press.

- James, W. (1977) [1904]. "The Chicago School," in *The Works of William James*, vol. 5. Cambridge, MA: Harvard University Press.
- James, W. (1979) [1897]. *The Will to Believe and Other Essays in Popular Philosophy*, ed. F. Burkhardt, F. Bowers, and I. K. Skrupselis, vol. 6 in F. Burkhardt (gen. ed.), *The Works of William James*. Cambridge, MA: Harvard University Press.
- Misak, C. (2016). *Cambridge Pragmatism: From Peirce and James to Ramsey and Wittgenstein*. Oxford: Oxford University Press.
- Peirce, C. S. Charles S. Peirce Papers, Houghton Library, Harvard University.
- Peirce, C. S., *The Collected Papers of Charles Sanders Peirce*, vols. 1–6 (1931–5), ed. C. Hartshorne and P. Weiss; vols. 7–8 (1958), ed. A. Burks. Cambridge, MA: Belknap Press.
- Peirce, C. S., *The Writings of Charles S. Peirce: A Chronological Edition*, gen. ed. N. Houser. Bloomington: Indiana University Press.
- Pratt, J. B. (1909) *What is Pragmatism?* Reprinted in vol. 1 of J. Shook (ed.), *Early Critics of Pragmatism* (2001), 5 vols. Bristol: Thoemmes Press.
- Quine, W. V. O. (1980 [1951]). "Two Dogmas of Empiricism," in his *From a Logical Point of View*, 2nd revised edn. Cambridge, MA: Harvard University Press, 20–46.
- Quine, W. V. O. (1981). "The Pragmatist's Place in Empiricism," in R. Mulvaney and P. Zeltner (eds.), *Pragmatism: Its Sources and Prospects*. Columbia: University of South Carolina Press, 21–39.
- Quine, W. V. O. (1987). *Quiddities: An Intermittently Philosophical Dictionary*. Cambridge, MA: Belknap Press.
- Quine, W. V. O. (2008 [1975]). "On Empirically Equivalent Systems of the World," in his *Confessions of a Confirmed Extensionalist and Other Essays*. Cambridge, MA: Harvard University Press, 228–43.
- Ramsey, F. (1990). *F. P. Ramsey: Philosophical Papers*, ed. D. H. Mellor. Cambridge: Cambridge University Press.
- Ramsey, F. P. (1990 [1926]). "Truth and Probability," in *F. P. Ramsey: Philosophical Papers*. ed. D. H. Mellor. Cambridge: Cambridge University Press, 52–94.
- Ramsey, F. (1991 [1930]). *On Truth*, ed. N. Rescher and U. Majer. Dordrecht: Kluwer.
- Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, R. (1982). *Consequences of Pragmatism (Essays 1972–80)*. Minneapolis: University of Minnesota Press.
- Rorty, R. (1991). *Objectivity, Relativism, and Truth (Philosophical Papers, vol. 1)* Cambridge: Cambridge University Press.
- Rorty, R. (1995). "Response to Richard Bernstein," in H. J. Saatkamp (ed.), *Rorty and Pragmatism: The Philosopher Responds to his Critics*. Nashville: Vanderbilt University Press, 68–71.
- Rorty, R. (2000). "Universality and Truth," in R. B. Brandom (ed.), *Rorty and his Critics*. Oxford: Oxford University Press, 1–30.
- Sellars, W. (1962). "Truth and Correspondence," *Journal of Philosophy* 59: 29–56.
- Skrupselis, I. K. and E. M. Berkeley (eds.) (1992–2004). *The Correspondence of William James*, 12 vols. Charlottesville: University Press of Virginia.
- Welchman, J. (1995). *Dewey's Ethical Theory*. Ithaca, NY: Cornell University Press.
- White, M. (1999). *A Philosopher's Story*. University Park, PA: University of Pennsylvania Press.
- Wiggins, D. (2002). "Marks of Truth: An Indefinibilist cum Normative View," in R. Shantz (ed.) *What is Truth?* Berlin: DeGruyter, 316–32.
- Wiggins, D. (2004). "Reflections on Inquiry and Truth Arising from Peirce's Method for the Fixation of Belief," in C. Misak (ed.), *The Cambridge Companion to Peirce*. Cambridge: Cambridge University Press, 87–126.

PART IV

TRUTH IN
METAPHYSICS

CHAPTER 12

PROPOSITIONS AND TRUTH-BEARERS

JEFFREY C. KING

THOSE who believe in propositions take them to play a number of roles in philosophy of language and related areas.¹ Propositions are thought to be the information contents of natural language sentences. Thus, sentences of different languages that have the same information content, such as “Snow is white” and “Schnee ist weiss,” are thought to *express the same proposition*. The proposition expressed by a sentence is thought to be (at least one of the things) asserted by a serious utterance of the sentence. When one understands a sentence, one grasps the proposition it expresses. The proposition expressed by a sentence is its meaning: it is what a compositional semantics assigns to the sentence. Further, propositions are primary bearers of truth-values. A true sentence is one that expresses a proposition that is true; a true belief in one whose propositional content is true. Propositions are also thought to be the bearers of modal attributes: they are possible, necessary, and impossible. They are also the things we doubt, believe, assume, and hope. Indeed, believing, doubting, and so on are often called *propositional attitudes* in virtue of the fact that many take them to be mental states the objects of which are propositions.² Further, propositions are thought to be designated by *that*-clauses such as “that snow is white.” Hence a sentence like “It is true that snow is white” is thought to predicate truth of the proposition that snow is white; a sentence like “Rebecca believes that snow is white” is thought to assert that Rebecca stands in the relation of belief to the proposition that snow is white; and a sentence like “It is possible that snow is white” is thought to predicate the attribute of *being possible* of the proposition that snow is white. Finally, any account of *context of utterance*, such as that of Stalnaker (1999), on which contexts include a collection of propositions knowingly commonly accepted

¹ It is not claimed that all proponents of propositions agree that they play all the roles listed in what follows. At least one does: the author of the present chapter. Also, for the moment contextual sensitivity is being ignored.

² For a recent view that denies this latter claim, see Moltmann (2014).

by conversational participants for the purposes of the conversation obviously requires propositions to characterize the notion of a *context of utterance*.

But what are these things, propositions, that many philosophers take to play all these roles? Most contemporary thought about propositions is traceable in one way or another to the classical views of propositions of the German mathematician and philosopher Gottlob Frege and the English logician and philosopher Bertrand Russell. Hence, we begin with their views.³

Frege held that linguistic expressions, whether syntactically simple or complex, had two entities associated with them by the rules of language: a *sense* and a *referent*. Frege does not give a lot of examples of the senses of different expressions, but he does at one point suggest that the sense of the proper name “Aristotle” could be *the student of Plato and teacher of Alexander the Great*. Frege held that simple expressions are conventionally attached to senses, which in turn determine the referents of the expressions. Aristotle himself is the referent of “Aristotle.” Putting these points together suggests that the sense of an expression is something like a descriptive condition that is uniquely satisfied by the referent of the expression. Thus, Aristotle is the referent of “Aristotle” in virtue of uniquely satisfying the descriptive condition *being the student of Plato and teacher of Alexander the Great* (assuming the sense of “Aristotle” is as suggested above). Of course, an expression may have a sense, but lack a reference because nothing satisfies the descriptive condition that is the sense. Presumably, “the greatest natural number” is an example. In understanding a linguistic expression, one *grasps* its sense.

Frege also held that there were two fundamentally different kinds of expression in natural languages. On the one hand, there are expressions like ordinary proper names, definite descriptions (expressions of the form *the F*—e.g. “the greatest natural number”) and sentences that are what Frege calls *complete* or *saturated*.⁴ Frege calls such expressions *proper names*, which is, then, something of a technical term for Frege. On the other hand, there are expressions like verbs, adjectives, and nouns that are *unsaturated* or *in need of supplementation*. Frege viewed the unsaturatedness of these expressions as their having “empty places.”⁵ Such unsaturated expressions can have different numbers of empty places. The transitive verb “loves” has two, whereas the intransitive “swims” has one. Frege calls unsaturated expressions with one argument place *concept expressions*; and those with more than one argument place *relation expressions*. One way to form a sentence is to take a concept or relation expression—e.g. the verb “swims”—and “fill” its empty places with appropriate proper names: “Rebecca swims.”

Frege thought that just as concept and relation expressions are unsaturated, so too are their senses and referents.⁶ A concept expression has a sense that is unsaturated

³ Because Frege and Russell scholarship has become so sophisticated and contested, the following characterizations should be taken to reflect how Frege and Russell strike many philosophers of language rather than as a scholarly account of their views.

⁴ Frege (1964: 36).

⁵ Frege (1891a: 139).

⁶ Frege (1891b: 174, n. B).

and in need of completion. Its referent is a function from objects to truth-values. Frege thought that functions are unsaturated or in need of supplementation because they take arguments. Frege calls (one-place) functions whose values are truth-values *concepts*. A relation expression also has an unsaturated sense and referent. For example, a relation word with two empty places expresses a doubly unsaturated sense and refers to a two-place function whose values are truth-values, which Frege calls a *relation*.⁷ By contrast, proper names—complete, saturated expressions—have complete senses and referents. Frege calls such complete referents *objects*.

Frege seems to have held principles of compositionality for both senses and referents.⁸ The sense of a complex expression is a function of the senses of its component words and how they are put together. Similarly, the reference of a complex expression is a function of the referents of the words in it and how they are combined. In the case of the senses of complex expressions, Frege seems to have held that they are literally built out of the senses of the words in the complex expressions in the way that the complex expressions themselves are built out of words.⁹

Having discussed the senses and referents of the words in simple “atomic” sentences like “Rebecca swims” and “Rebecca loves Carl,” we now turn to the senses and referents of such sentences. Frege thought that quite generally, “complex senses,” senses with other senses as parts, are the result of one or more senses “completing” or “saturating” an unsaturated sense.¹⁰ In the case of atomic sentences such as those above, we have the (singly or doubly) unsaturated sense of the verb being completed by the complete proper name senses. The result is once again a complete sense. Frege calls the complete senses of declarative sentences *thoughts*. Thoughts are true or false; thoughts are what are grasped when one understands a sentence. Hence, Frege’s thoughts are what other philosophers call propositions. Frege famously held that the referents of sentences are truth-values. Since sentences are complete expressions, proper names, they must refer to objects.¹¹ Thus, the Fregean doctrine that the True and the False are objects.

As for thoughts expressed by non-atomic sentences, connectives like “not” and “and” for Frege express singly and doubly unsaturated senses, respectively. When you complete the unsaturated sense of “not” with a thought, you once again get a thought: the negation of the original thought. Similarly, when you complete the doubly unsaturated sense of “and” with two thoughts, the result is again a thought (their conjunction). Thus, the thought expressed by “It is not the case that snow is white” is a thought consisting of the thought that snow is white completing the unsaturated sense of “It is not the case.”

⁷ Frege (1891a: 146).

⁸ Frege (1892a: 156–7).

⁹ Frege (1919: 364–5).

¹⁰ Frege (1892b: 193); Frege (1919: 363).

¹¹ Of course, just as Frege thought some ordinary proper names lack referents (e.g. “Odysseus”), so he thought some sentences lack referents (e.g. those containing ordinary proper names without referents). A sentence like “Odysseus was strong” expresses a thought, but the thought does not determine any referent, and so is neither true nor false.

Similar remarks apply to the thoughts expressed by conjunctive sentences. Connectives like these refer to one- and two-place functions from truth-values to truth-values. Hence, the connectives refer to concepts and relations (this is the special case where the objects that are the *arguments* of the concepts and relations are themselves truth-values).

Finally, determiners like “every,” “some,” “no,” etc. express doubly unsaturated senses.¹² However, to yield a thought, these senses must be combined with the senses of two (first-level) concept words. Thus, the thought expressed by “Every dog barks” is the result of completing the doubly unsaturated sense of “every” with the senses of the (first-level) concept expressions “dog” and “barks” (in that order). On the other hand, “every” refers to a function from a pair of first-level concepts (functions from objects to truth values) to truth-values. As such, the referent of “every” is a *second-level relation* (in virtue of taking first-level concepts as arguments and yielding truth-values as values). The relation “every” refers to maps concepts A and B (in that order) to true iff every object that A maps to true, B maps to true as well.

We have now seen what the senses and references are of proper names, predicates (first-level concept and relation expressions), truth functional connectives, and determiner phrases. We have also seen what sorts of thought are expressed by “atomic” sentences, sentences containing truth functional connectives, and quantified sentences. We now turn to general features of Frege’s view of thoughts.¹³

Obviously, thoughts are composed out of other senses, at least one of which must be unsaturated. Further Frege was quite explicit that thoughts exist timelessly and independently of any thinker. True thoughts are true prior to and independently of any thinker thinking them or recognizing their truth.¹⁴ Frege famously held that thoughts are not part of the external physical world, nor are they things, like subjective experiences, that belong to the consciousness of a thinker. They are non-spatial and non-temporal. They occupy, Frege claimed, a third realm. Of course, this means that the senses thoughts are constructed out of exist timelessly and independently of thinkers, and occupy this third realm as well.

There are two final points to stress about Frege’s view, since they are relevant to issues that will be discussed later. First, he seems to have held that thoughts are things that by their very natures and independently of mind and languages are true or false. Second, he thought he had a sort of explanation of how the constituents of thoughts—senses of subsentential expressions—are held together. It is a matter of senses “saturating” unsaturated senses. However, since Frege himself regarded talk of unsaturated senses as a metaphor¹⁵ and since he really gives no account of what unsaturatedness is, his account of what holds thoughts together seems more of a promissory note than a substantive account.

¹² Frege (1892b: 187).

¹³ Frege’s term “thoughts” will continue to be used, but the reader should remember that what is being discussed is Frege’s view of what other philosophers would call “propositions.”

¹⁴ Frege (1918: 337).

¹⁵ Frege (1892b: 193).

Despite the seminal importance of Frege's views on referents, thoughts, and senses generally, it is fair to say that few contemporary philosophers are attracted to the idea of Fregean senses, thoughts included, at least as Frege construed them. Many have found the idea of these timelessly existing, mind- and language-independent occupiers of the third realm obscure. But the waning popularity of Fregean senses is most attributable to more direct arguments that they cannot do the theoretical work required of them. On the one hand, Saul Kripke (1980) gave powerful arguments against the view that ordinary proper names have anything like Fregean senses. Similarly, David Kaplan (1977) argued compellingly that indexicals ("I," "here," "yesterday") and demonstratives ("he," "that") were most amenable to a non-Fregean treatment. Once it appeared that a significant number of types of words were not best treated as having senses, the idea of any words having Fregean senses became much less attractive. As will be discussed below, by the mid-1980s, many philosophers found themselves attracted to a more Russellian view of propositions. We turn now to Russell's view of propositions.

In *The Principles of Mathematics*, Bertrand Russell outlined what was to become a very influential account of propositions. Russell uses the expression "term" for anything that is a constituent of a proposition. Since Russell says that a chimera is a term, terms apparently include things that in any ordinary sense don't exist. There is a fundamental distinction among terms between *things* and *concepts* for Russell. Things are *indicated* by proper names and concepts are indicated by all other words.¹⁶ Russell had the idea that for many propositions, we can distinguish between the *subject* and the *assertion about the subject*. In simple "subject-predicate" propositions, such as that expressed by "Socrates is human," there is only one way of analyzing the proposition into subject (Socrates) and assertion (being human). In the case of "relational propositions" (that expressed by "A is larger than B"), there is more than one way of analyzing the proposition into subject and assertion: either A or B may be considered the subject. For Russell, the *terms of a proposition* (as opposed to terms *simpliciter*) are the terms in the proposition that can be considered subjects. Thus, in the proposition expressed by "Socrates is human," Socrates is the term of the proposition; in that expressed by "A is larger than B," A and B are both terms of the proposition. Now Russell thought that concepts can occur in two ways in propositions: they can either occur as terms of the proposition ("Humanity belongs to Socrates"—humanity is a term of this proposition) or not ("Socrates is human"). It is characteristic of things, as opposed to concepts, that they can only occur in propositions as terms of propositions. Among concepts, Russell distinguishes between *predicates*, which are indicated by adjectives, and *relations*, which are indicated by verbs.

Russell seems to have thought that relations were the key to explaining two central features of propositions. On the one hand, propositions, like that expressed by "Caesar died," assert something, unlike what is expressed by "the death of Caesar." Yet the words

¹⁶ "Indication" is the word Russell uses for the relation between a word and the term it contributes to propositions expressed by sentences in which it occurs.

in both linguistic strings seem to indicate the same terms (Caesar, and the concept *dying*).¹⁷ Russell thought that when a verb is used as a verb, as in “Caesar died,” we get a proposition in which something is asserted and so the proposition is capable of being true or false. However, when the verb is made into what Russell calls a *verbal noun*, as in “the death of Caesar,” we cease to have an assertion and so cease to have something capable of being true or false. Hence Russell thought that a verb being used *as a verb* in some sense is the key to why propositions assert something and so are true or false. Only when the relation indicated by the verb occurs in the proposition *in a certain way* do we get something asserted and an entity capable of being true or false.

On the other hand, and related to these points, Russell was puzzled by the difference between the proposition that A differs from B and what he considered the analysis of the proposition, which consisted of simply specifying its constituents: A, difference, B.¹⁸ Because the analysis is not the proposition, we must explain what distinguishes the proposition from the mere sum of its constituents. The problem of providing such an explanation is called by many *the problem of the unity of the proposition*.¹⁹ Russell seems to have thought that relations, indicated by verbs, provided the explanation of the difference between a proposition and a mere sum of its constituents. In the proposition that A differs from B, the relation of difference actually relates A and B. What Russell seems to mean by this is that the proposition consists of A and B standing in the difference relation (in that order).²⁰ However, in the mere sum of the proposition’s constituents—A, difference, B—the difference relation is not relating A and B. Thus Russell writes:

Consider, for example, the proposition “A differs from B.” The constituents of this proposition, if we analyze it, appear to be only A, difference, B. Yet these constituents, thus placed side by side, do not reconstitute the proposition. The difference which occurs in the proposition actually relates A and B, whereas the difference after analysis is a notion which has no connection with A and B. It may be said that we ought, in the analysis, to mention the relations which difference has to A and B, relations which are expressed by *is* and *from* when we say “A is different from B.” These relations consist in the fact that A is referent and B relatum with respect to difference. But “A, referent, difference, relatum, B” is still merely a list of terms, not a proposition. A proposition, in fact, is essentially a unity, and when analysis has destroyed the unity, no enumeration of constituents

¹⁷ How Russell could have held that all verbs express relations, by which he seems to have meant two-place relations, in Russell (1938 [1903]), is a matter of some obscurity.

¹⁸ This seems to be a bad way to think of the analysis of a proposition, since it would seem to entail that distinct propositions have the same analysis. If the proposition that John loves Sue and the proposition that Sue loves John have the same constituents, as would seem to be the case for Russell, then both propositions have as their analysis: John, loving, Sue.

¹⁹ See King (2009) for an argument that there is no single problem that uniquely deserves the epithet *the problem of the unity of the proposition*.

²⁰ Here the relation indicated by the verb is symmetrical, and so order isn’t crucial. But of course when we move to things like the proposition that A is larger than B, order becomes crucial.

will restore the proposition. The verb, when used as a verb, embodies the unity of the proposition, and is thus distinguishable from the verb considered as a term, though I do not know how to give a clear account of the precise nature of the distinction.²¹

An obvious problem with such an account of the unity of propositions is that it is unclear how to account for false propositions. Consider the proposition that A is larger than B. This is to be the entity consisting of A standing in the *larger than* relation to B. But one would think that if A stands in the *larger than* relation to B, A must in fact be larger than B, so that the proposition is true. But if A fails to be larger than B, then A does not stand in the *larger than* relation to B. But the proposition that A is larger than B was just supposed to be A standing in that relation to B. So it appears that when A is not larger than B, there is nothing that is the proposition that A is larger than B. These considerations appear to be related to Russell's own reasons for rejecting his 1903 theory of propositions, and propositions generally, by 1910.²²

In any case, it is worth summarizing the features of Russell's account of propositions that proved significant historically.²³ Russell held that individuals like Socrates were constituents of propositions such as the proposition that Socrates was human. Similarly, the *taller than* relation is a constituent of the proposition that Aristotle was taller than Plato. So for Russell, individuals and relations were constituents of propositions, where these elements were bound together in the proposition, presumably yielding a complex structured entity. As we'll see, many contemporary philosophers follow Russell in thinking of propositions in this way.

Developments in modal logic in the late 1950s and early 1960s led philosophers to think of propositions in a way very different from the way Russell and Frege thought of them. Though an exposition of modal logic is beyond the scope of the present work, presentation of some of the basic ideas of model theory for propositional modal logic will be helpful. Some familiarity with propositional logic is presupposed.

Let, P, Q, and R with or without numerical subscripts be atomic formulae. Complex formulae are built up with \sim , $\&$, and \Box as follows: if A and B are formulae, so are $\sim A$, $A\&B$, and $\Box A$. $\Box A$ should be thought of as *necessarily A*. We want to formally implement the idea that to be necessary is to be true no matter how the world happens to be: to be true in all possible worlds. Hence, for $\Box A$ to be true—true in the actual world—A must be true in all *possible worlds*. For our semantics, we begin with a pair $\langle W, @ \rangle$ where W is a set and $@ \in W$. Intuitively, W is the set of all possible worlds and @ is the actual

²¹ Russell (1938 [1903]: 49–50).

²² In Russell (1994 [1910]).

²³ I have only considered “atomic proposition” in discussing Russell's view mainly because it is the features of his account of such propositions that ended up having the most influence. Russell abandoned his 1903 account of “quantified propositions” by 1905 and the 1903 account is currently of only historical interest.

world. Next, we introduce a function f such that for any atomic formula A and any $w \in W$, $f(A, w) = T$ or F . We extend the domain of f to all formulae in the usual way (let A, B be formulae):

1. $f(\sim A, w) = T$ if $f(A, w) = F$; otherwise $f(\sim A, w) = F$.
2. $f(A \& B, w) = T$ if $f(A, w) = f(B, w) = T$; otherwise $f(A \& B, w) = F$.
3. $f(\Box A, w) = T$ if for all $w' \in W$, $f(A, w') = T$; otherwise $f(\Box A, w) = F$.²⁴

Clause 3 captures the idea that to be necessarily true is to be true in all possible worlds. We'll call a pair $\langle W, @, f \rangle$ a *model* for our language. Obviously, as f 's initial domain and clauses 1–3 make clear, in a model, formulae of our language have truth-values only relative to worlds (members of W). So given a model M , and any formula A , we can define a function $g_{A,M}$ whose domain is W as follows: for any $w \in W$, $g_{A,M}(w) = T$ if A is T at w relative to M ; otherwise, $g_{A,M}(w) = F$. Following established usage, we can call $g_{A,M}$ the *intension of A relative to M* . Once intensions are in the picture, it is natural to think of \Box as an operator on intensions: \Box "looks at" the intension of the formula it embeds. If that intension maps every $w \in W$ to T , the \Box -formula is T relative to M at the world where it is being evaluated. This in turn makes it natural to think of the intension as being the thing with modal attributes. An intension I is *necessary* at a world w iff for all w' , $I(w') = T$. Recall that propositions were supposed to be the things that possess modal attributes. But then from this perspective, it becomes natural to identify the proposition expressed by a formula A in a model M with the intension of A relative to M . Thus, it is natural to view propositions as functions from possible worlds to truth-values (or as the sets of worlds whose characteristic functions they are).

The sort of approach to the semantics of modal logic outlined above is often called *possible worlds semantics*. Impressed by the successes of possible worlds semantics in modal logic, logically-minded philosophers of language began to formulate possible world semantics for natural languages.

Prominent examples include Montague (1968) and Lewis (1970), though both of these works represent a generalization of possible worlds semantics. Instead of having sentence intensions be functions from worlds to truth-values, Montague's (1968) sentence intentions are functions from worlds and times to truth-values, whereas Lewis's (1970) are functions from worlds, contexts, times, places, and so on to truth-values. These arguments for sentence intensions were typically represented as tuples. So on Montague's view we would have a pair of a world and time: $\langle w, t \rangle$, whereas for Lewis we would have a bigger tuple of a world, context, time, place, and so on. Such tuples were usually called *indices* and these semantic approaches generalizing possible world semantics were sometimes called *index semantics*. Having said this, we'll restrict ourselves to possible world semantics here.

²⁴ Here I suppress the fact that one often introduces an accessibility relation R on the set W and writes clause 3 as follows: $f(\Box A, w) = T$ if for all $w' \in W$ such that wRw' , $f(A, w') = T$; otherwise $f(\Box A, w) = F$.

On such an approach to the semantics of natural language, expressions of the natural language get assigned *extensions* relative to possible worlds, just as in our modal logic formulae get assigned truth values—extensions of formulae—relative to possible worlds. Thus, names, n -place predicates, and sentences will be assigned individuals, sets of n -tuples of individuals, and truth-values, respectively, as extensions relative to possible worlds. For many expressions, we should expect their extensions to vary from world to world. Take the predicate “dog.” Given that different possible worlds contain different dogs, the extension of “dog”—the set of things that are dogs at the world in question—will vary from world to world. But it is natural to think of such expressions as having meanings that determine their extensions at different worlds. Such a meaning must determine a function from worlds to the expression’s extensions at worlds. For sentences, these functions would be functions from worlds to truth-values. On analogy with the functions from worlds to truth-values associated with formulae (relative to a model) discussed above, functions from worlds to extensions associated with expressions are called *intensions*.

Above it was said that meanings of expressions must determine intensions. The simplest way meanings could do this is by *being* intensions. Thus, within the framework of possible world semantics as applied to natural language, it is very natural to identify an expression’s meaning with its intension. Similarly, we know that the proposition expressed by a sentence must determine its intension. Again, the simplest way it could do this is by being the intension. Thus, we arrive at the view that propositions are sentence intensions: functions from possible worlds to truth-values, or sets of possible worlds. Robert Stalnaker (1987; 1999; 2003) is the best-known defender of this view.

Such a view individuates propositions in a very coarse-grained manner and many have found this feature of the view a reason to reject it.²⁵ It entails that there is only one necessary proposition (and one impossible proposition), since there is only one function that maps every world to true (and only one set of all possible worlds). Among other things, this means that there is only one mathematical truth (assuming mathematical truths are necessarily true). Further, if we assume that *that*-clauses designate propositions, and that “Stella believes that snow is white” asserts that Stella stands in the belief relation to the proposition that snow is white, this means that any time two sentences are true in all the same possible worlds, the result of embedding them under something of the form “S believes” cannot diverge in truth value. But *prima facie* this seems false: “Stella believes that $2 + 2 = 4$ ” might be true while “Stella believes that arithmetic is incomplete” is false.

Despite these consequences of the view that propositions are functions from possible worlds to truth-values, which were appreciated early on, the view was the dominant view of propositions for many years. Nonetheless, many philosophers were eager to find an account of propositions on which they were individuated more finely than on the possible worlds approach. In the 1980s and 1990s, various such accounts were

²⁵ Though Lewis (1970) defines more fine-grained sentence meanings.

developed. We will discuss two significant accounts of the sort, which we will call *the algebraic account* and *the neo-Russellian account*, respectively.²⁶

Algebraic accounts were developed in Bealer (1982; 1993; 1998); Menzel (1993); and Zalta (1988).²⁷ There are important differences in details between these accounts, but they are sufficiently similar in motivation and content to be usefully grouped together under the heading *algebraic accounts*. Here we shall focus on the presentation in Bealer (1993a; 1998). Bealer begins with the idea that properties, relations, and propositions are *sui generis* entities not reducible to anything else. That is, they are primitive and irreducible. There are then fundamental logical operations on these things: *conjunction* (the proposition A&B is the conjunction of the proposition that A and the proposition that B), *singular predication* (the proposition that o is F is the singular predication of the property F of o), *existential generalization* (the proposition that there exists an F is the existential generalization of the property F) and so on. One then analyzes the behavior of the fundamental *sui generis* entities—properties, relations, and propositions—by constructing what Bealer [1998] calls *intensional model structures* (IMS). An IMS is a triple $\langle D, \tau, K \rangle$, where D is a domain that partitions into subdomains, D_{-1} , D_o , D_1 , $D_2 \dots$, where D_{-1} consists of particulars; D_o , propositions; D_1 , properties; D_2 , two-place relations; and so on. τ is a set of logical operations as described above; and K is a set of extensionalization functions. Where H is an extensionalization function, it assigns appropriate extensions to each item in D: truth-values to propositions (members of D_o); sets of items in the domain to properties; sets of ordered pairs to two-place relations, and so on.²⁸ There is a distinguished extensionalization function $G \in K$ that is the *actual* extensionalization function: it assigns members of D their actual extensions. Extensionalization functions are constrained to respect the logical operations in the obvious way: $H(\text{neg } p) = T$ iff $H(p) = F$ and so on. What makes an IMS *intensional* is that for some $x, y \in D_i \subset D$, $i \geq 0$, and some $H \in K$, $H(x) = H(y)$, but $x \neq y$.

Though strictly, for Bealer, propositions have no mereological parts and are primitive and simple, he does have a notion of *propositional constituency*. Each proposition has what we might call a *composition tree* indicating how it results from the application of logical operations on members of D.²⁹ For example, Figure 12.1 is such a tree for the proposition that o is F:

²⁶ Of course there have been many other fine-grained accounts of propositions developed since the 1980s (e.g. Schiffer 2003), but I consider the algebraic accounts and neo-Russellian accounts to be the most influential.

²⁷ Though considerations of fineness of grain seem to be the primary motivation for those who accept the algebraic accounts (e.g. see Menzel 1993: 62–4; Bealer 1993a: 20; and Bealer 1998: 5), other worries with possible worlds accounts are mentioned as well. e.g. Bealer (1993a: 20; 1998: 4) and Menzel (1993: 62) worry that possible worlds accounts commit one to denying actualism. However, even the most ardent supporter of the possible worlds account—Robert Stalnaker—has long held that the account is consistent with actualism. See Stalnaker (1987: ch. 3 (reprinted in Stalnaker 2003)) and Stalnaker (2003: 6).

²⁸ Members of D_{-1} are assigned themselves.

²⁹ For reasons that elude me, Bealer (1993a) actually defines a *decomposition tree* for propositions that shows how they can be decomposed into elements of D and the *inverses* of logical operations.

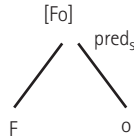


FIGURE 12.1

This shows that the proposition that *F* is *o* results from singularly predicating the property *F* of *o*. Now an item *x* is a constituent of a proposition *P* iff *x* appears somewhere in *P*'s composition tree. Hence, the property *F* and particular *o* are constituents of the proposition that *o* is *F* as desired. But to repeat, they are in no sense *parts* of the proposition. Though it will not be discussed here, this is one of the keys to the fact that Bealer's algebraic approach to propositions is consistent with the claim that a proposition may exist even if some of its constituents don't.³⁰

It should already be clear that on the algebraic approach, propositions are more fine-grained than on the possible worlds account. For example, the necessary proposition that *p* or not *p* ≠ the necessary proposition that not (*p* and not *p*). After all, these propositions have different constituents: the former has the logical operation of disjunction as a constituent and the latter doesn't. However, nothing so far gives us the materials for distinguishing the proposition that *Hesperus* is a planet from the proposition that *Phosphorus* is a planet. Bealer certainly wants his theory to achieve this. He does this by introducing what he calls *non-Platonic modes of presentation* associated with proper names.³¹ These might be the name itself, understood not as a mere orthographic entity but as fine-grained entity of some sort individuated by the practice of using it. Since the practice of using "Cicero" to refer to a town is different from the practice of using it to refer to a Roman orator, these are different names. Or they might be the naming practices associated with names. Or they might be *causal naming chains* associated with names. Whichever of these turns out to be the best option will be non-Platonic modes of presentation associated with names. By invoking these non-Platonic modes of presentation, Bealer is able to distinguish the proposition that *Hesperus* is a planet from the proposition that *Phosphorus* is a planet.

The second sort of fine-grained account of propositions developed in the 1980s and '90s was the *neo-Russellian* or *structured* account. Worries about the coarseness of grain of propositions on the possible worlds account certainly was one motivation for formulating neo-Russellian accounts. But other developments pointed in the direction of neo-Russellian accounts as well.

First, as mentioned above, Saul Kripke (1980) gave powerful arguments against the view that ordinary proper names are associated with descriptive conditions the satisfaction of which determine the bearers of the names. Kripke also defended the view that

³⁰ See Bealer (1998: 12–14).

³¹ He thinks of the elements of subdomain D_1 —properties—as *Platonic* modes of presentation.

ordinary proper names are *rigid designators*: they refer to their actual bearers in all possible worlds. To take Kripke's example, when we consider the sentence "Aristotle was fond of dogs" and evaluate it for truth and falsity at various possible worlds, it is always *Aristotle's* properties at the worlds in question that determine whether the sentence is true or false at that world. This is because the referent of "Aristotle" at each world is Aristotle. David Kaplan (1977) wanted to defend the view that indexicals and demonstratives are devices of *direct reference* (he thought the same about names but didn't really *argue* for that claim in Kaplan 1977); and he wanted to distinguish directly referential expressions from rigid designators. Kaplan sharply distinguished between *contexts* (*of utterance*) and *circumstances* (*of evaluation*). When a sentence like "I am hungry" is uttered in a *context*, a proposition is expressed. We can then ask whether that proposition is true or false at various *circumstances* (possible worlds).³² Kaplan's idea was that directly referential expressions may (indexicals and demonstratives) or may not (names) be associated with semantical rules that determine their referents in context. But once their referents are secured in context, those referents are all that is provided by the semantical rules of the expressions in question. Nothing is provided that could be evaluated at a circumstance. Take the word "I." It has a semantic rule that gives uses of "I" referents in contexts (a use of "I" in a context *c* refers to the speaker in *c*). But once that referent is secured, the rules do not yield anything that can be evaluated at a circumstance. They give you only the referent in the actual world, which is then taken to be the referent of the expression relative to that context in all circumstances. Kaplan writes,

For me, the intuitive idea is not that of an expression which *turns out* to designate the same object in all possible circumstances, but an expression whose semantical *rules* provide *directly* that the referent in all possible circumstances is fixed to be the actual referent. In typical cases the semantical rules will do this only implicitly, by providing a way of determining the *actual* referent and no way of determining any other propositional component.³³

Thus all directly referential expressions are rigid designators. But we can see that an expression can be rigid without being directly referential. An example would be "the successor of 2." This expression when used in a context yields a descriptive condition, which can then be evaluated at a circumstance. At any circumstance at which that descriptive condition is evaluated, we get the number 3. So the expression is rigid (designating 3 at every circumstance), but it is not directly referential, since when it is used in a context we get not a referent, but a descriptive condition that can be evaluated at different circumstances.

The crucial point for present purposes is that in a possible worlds framework, one cannot distinguish between directly referential designators and rigid designators. Singular terms like indexicals, names, and definite descriptions will be assigned intensions by the semantics (relative to contexts in the case of indexicals). If the

³² Actually, Kaplan's circumstances were world/time pairs, but I'll ignore that here.

³³ Kaplan (1977: 493).

expressions are rigid designators, the expressions will be assigned intensions that are constant functions, mapping every world to the same individual. However, since all the semantics assigns to singular terms are such intensions, there is no distinction between rigid expressions and directly referential expressions. Either an expression is assigned a constant function as intension or it isn't. There is nothing more to it. As a result, Kaplan invoked structured propositions with individuals, properties and relations as constituents, consciously echoing Russell's 1903 view, to distinguish directly referential expressions from merely rigid ones:

If I may wax metaphysical in order to fix an image, let us think of the vehicles of evaluation—the what-is-said in a given context—as propositions. Don't think of propositions as sets of possible worlds but, rather, as structured entities looking something like the sentences which express them. For each occurrence of a singular term in a sentence there will be a corresponding constituent in the proposition expressed. The constituent of the proposition determines, for each circumstance of evaluation, the object relevant to evaluating the proposition in that circumstance. In general, the constituent of the proposition will be some sort of complex, constructed from various attributes by logical composition. But in the case of a singular term which is directly referential, the constituent of the proposition is just the object itself. Thus it is that it does not just *turn out* that the constituent determines the same object in every circumstance, the constituent (corresponding to a rigid designator) just *is* the object. *There is no determining to do at all.*³⁴

So a directly referential expression contributes to the Russellian proposition (relative to a context) its referent/object (in that context). That is why, when we evaluate the proposition at different circumstances, that same referent is always relevant. It is built into the proposition being evaluated. By contrast, a merely rigid (i.e. rigid and not directly referential) expression contributes to propositions some complex that determines a descriptive condition. When the proposition is evaluated at any circumstance, the descriptive condition *turns out* to pick out the same object.

As more philosophers became attracted to the idea that indexicals, demonstrative pronouns, and names were directly referential and that this amounted to more than mere rigidity, they naturally were drawn toward a view of propositions—a neo-Russellian view of propositions as structured entities with individuals, properties, and relations as constituents—on which the distinction between merely rigid and directly referential expressions could be made, and away from the possible worlds view of propositions on which the distinction is obscured.³⁵

³⁴ Kaplan (1977: 494). Oddly in the penultimate sentence here, Kaplan talks about the constituent corresponding to a rigid designator, where he presumably means the constituent corresponding to a directly referential term.

³⁵ Despite using Russell's view of propositions to explain the difference between merely rigid expressions and directly referential expressions, Kaplan explicitly said that this view of propositions was not part of his official theory (1977: 496); and his formalism employs possible worlds semantics.

A second factor that drove philosophers away from a possible worlds account and toward a neo-Russellian view was that many people had become increasingly uncomfortable with the coarse-grained way that the former theory individuated propositions and the consequences of that discussed above. Further, Scott Soames (1987) produced powerful reasons for thinking that any account of propositions on which they were sets of metaphysically possible worlds (or indeed any sort of truth supporting circumstances of evaluation) was fatally flawed.

As a result of these two factors, by the late 1980s and early 1990s, many philosophers had abandoned the possible worlds account of propositions in favor of a neo-Russellian or structured approach. Prominent defenders of such views include Scott Soames (1987), Nathan Salmon (1986), and Mark Richard (1990).

Though many philosophers had adopted a neo-Russellian account of propositions on which they were structured entities with individuals, properties and relations as constituents, Jeffrey C. King (2007) pointed out that surprisingly little was being said about what structured propositions actually were. Structured propositions were often represented without further comment by their advocates as *n*-tuples of objects, properties and relations.³⁶ The proposition that Stella is happy was represented as *<Stella, being happy>*. But of course as King (2007) pointed out, *representing* propositions as *n*-tuples is not to say anything about what propositions actually *are*. King highlighted the fact that structured proposition theorists were in the unsatisfactory position of saying propositions are structured but not saying anything about what they are.

Perhaps some of the advocates of structured propositions even intended to *identify* propositions with such *n*-tuples. However, King (2007) points out that there are two problems with this suggestion. First, there are many different *n*-tuples that all seem to be equally good candidates for being some propositions. Consider the proposition that Rebecca loves Carl. All of the following tuples seem equally good candidates for being that proposition:

<Rebecca, loving, Carl> *<loving, Rebecca, Carl>* *<Rebecca, Carl, loving>*

<Rebecca, <loving, Carl>> *<<Rebecca, loving>, Carl>* *<<Rebecca, Carl>, loving>*

and so on. This suggests that the proposition is not any of these *n*-tuples. Second, and more importantly, King pointed out that it doesn't seem that this view could provide any explanation of why propositions have truth conditions. Many *n*-tuples do not have truth conditions. So why do the *n*-tuples that are propositions have truth conditions? It certainly does not seem as though an acceptable answer to this question is in the offing.

We now turn to recent theories of propositions. Limitations of space require that only a brief overview will be given of each theory discussed and that there will be no attempt

³⁶ Salmon (1986: 25 and Appendix C); Soames (1987: 222, 224–6).

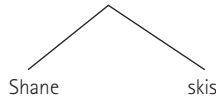


FIGURE 12.2

to critically assess the theories. The hope is to give the reader a sense of what theories are currently available so that she may investigate them further.

As indicated above, Jeffrey C. King (2007) noted that structured proposition theorists had given no real account of what structured propositions are. King (2007) was the first attempt in the recent literature to formulate a substantial theory of what they are.³⁷ It is worth highlighting three novel ideas in King (2007) and in subsequent work (King 2009; 2011; 2013; King, Soames, and Speaks 2014) since they were subsequently picked up by Scott Soames and Peter Hanks, who are discussed below. First, King rejected as mysterious theories according to which propositions have truth conditions and so represent the world as being a certain way by their very natures and independently of minds and languages. He claims that no one has any idea how something could have truth conditions by its very nature and independently of minds and languages. Second, and closely related to this, he claimed that an adequate theory of propositions must *explain* how/why propositions have truth conditions and so represent the world as being a certain way.³⁸ Third, since King claims that representational capacities of propositions cannot be something they have inherently and by their natures, he holds that their representational capacities must derive from and be explained by the representational capacities of thinking agents.

King claims that propositions consist of their constituents standing in a relation that he calls *the propositional relation*.³⁹ So, for example, the proposition that Shane skis consists of Shane standing in the propositional relation to the property of skiing. To explain what the propositional relation is, let's begin by looking at the simple sentence "Shane skis" and its syntactic structure (Figure 12.2).⁴⁰

Call the syntactic relation that obtains between "Shane" and "skis" in the sentence here "R." King calls relations like R that lexical items stand in to form sentences *sentential relations*. King points out that English speakers *interpret* R in a certain way: they take

³⁷ This is noted by Jeff Speaks in the Introduction to King, Soames, and Speaks (2014).

³⁸ King claimed as early as the 1990s that explaining how/why propositions had truth conditions was an important desideratum for a theory of propositions. See e.g. King (1995).

³⁹ We'll discuss King's most recent formulation of his view in King, Soames, and Speaks (2014). King (forthcoming) formulates a new version of his view that individuates propositions less finely than the latter.

⁴⁰ In using this sentence as an example, King is idealizing a lot about its syntactic structure. John Collins (2007) has recently argued that King's pretending syntax is much simpler than it is for expository purposes is far from innocent, since the real complexity of syntax ends up being a problem for him. King (2011) responds to Collins.

R to *ascribe* the semantic value of “skis” to the semantic value of “Shane.” This is in part why the English sentence is true iff Shane possesses the property of skiing.

As to what is meant by talk of *interpreting* R here, King claims that English speakers interpreting R as ascribing the semantic value of “skis” to the semantic value of “Shane” consists in the fact that they spontaneously and unreflectively take 1 to be true iff Shane possesses the property of skiing. Similar things happen when English speakers confront other instances of syntactically concatenated expressions: they spontaneously and unreflectively compose the semantic values of the concatenated expressions in characteristic ways. For example, when English speakers confront “grey house” they do something like conjoin the properties expressed by “grey” and “house.” In general, that speakers *interpret* syntactic concatenation in the ways they do consists in the fact that they spontaneously and unreflectively compose the semantic values of the concatenated expressions in a small handful of ways, including the ways described. King puts the fact that English speakers interpret R as ascribing the semantic value of “skis” to the semantic value of “Shane” by saying that *R encodes ascription in English*.

King notes that it appears that speakers of different natural languages interpret syntactic concatenation in *the same* small handful of ways. He holds that this is so because it is part of our biologically-endowed language faculty. Since the claim that speakers interpret syntax in certain ways will play a role in the explanation of how propositions end up having truth conditions for King, it is important that their interpreting syntax not involve their having propositional attitudes. He thinks that the fact that their interpreting syntax in the ways they do—composing semantic values of concatenated expressions in the ways they do—is hard-wired as a result of our biologically-endowed language faculty gets around this worry.⁴¹

Now in virtue of the existence of the English sentence 1 or its Spanish translation, there is a two-place relation that Shane stands in to the property of skiing: *there is a context c and assignment f such that ____ is the semantic value relative to c and f of a lexical item e of some language L and ____ is the semantic value relative to c and f of a lexical item e' of L such that e occurs at the left terminal node of the sentential relation R that in L encodes ascription and e' occurs at R 's right terminal node*. This relation, King claims, is the relation that holds Shane and the property of skiing together in the proposition that Shane skis: it is the propositional relation of that proposition.

King (2007; 2009) calls an object possessing a property, or n objects standing in an n -place relation, or n properties standing in an n -place relation and so on a *fact*. Then, according to King, the proposition that Shane skis is the fact consisting of Shane and the property of skiing standing in the two-place propositional relation mentioned above.

Now that King has said what structured propositions are, he goes on to explain how/why they have truth conditions. For the moment, let's suppress the fact that for King the propositional relation is complex (e.g. it has the sentential relation R of Figure 12.2 as a component or “part”), and concentrate on the idea that on his view the proposition that

⁴¹ This is discussed in detail in King (forthcoming).

Shane skis is a fact consisting of Shane standing in the two-place propositional relation to the property of skiing. Call this proposition *P* and picture it in tree from like Figure 12.2 with the branches representing the propositional relation. Now if speakers interpreted the propositional relation of *P* as *ascribing* the property of skiing at its right terminal node to Shane at its left terminal node, then the fact would be true iff Shane possessed the property of skiing. Recall that King claims that the *sentential* relation of the *sentence* of Figure 12.2 is interpreted by English speakers as ascribing the property that is the semantic value of “skis” to the semantic value of “Shane,” which we expressed by saying that the sentential relation *R* encodes ascription in English. Recall too that King claims that speakers are hard-wired to do this by their biologically-endowed language faculty. King argues that interpreting the *propositional* relation of *P* as ascribing the property of skiing to Shane *just is the same thing* as interpreting the *sentential* relation of Figure 12.2 as encoding ascription.⁴² Hence in the end, the explanation for propositions having truth conditions for King is traceable to speakers interpreting syntax and hence ultimately to our biologically-endowed language faculty.

We now turn to the theory of propositions formulated by Scott Soames (2010; 2015) and King, Soames, and Speaks (2014). As mentioned above, Soames follows King in endorsing the three novel claims mentioned above from King (2007) and subsequent work: Soames (i) rejects theories according to which propositions have truth conditions and so represent the world as being a certain way by their very natures and independently of minds and languages as mysterious; (ii) holds that an adequate theory of propositions must *explain* how/why propositions have truth conditions and so represent the world as being a certain way; and (iii) holds that since the representational capacities of propositions cannot be something they have inherently and by their natures, their representational capacities must derive from and be explained by the representational capacities of thinking agents. However, Soames’s positive account of how/why propositions have truth conditions, and his theory of propositions more generally, differs in important ways from King’s.

Soames begins with the notion of the mental act of *predication*, which he takes to be primitive. However, he provides the following examples. If an agent perceives an object *o* as red, and so has a perceptual experience that represents *o* as being red, the agent *predicates* redness of *o*. If an agent thinks of *o* as red, she thereby predicates redness of *o*.⁴³ For Soames, predicating redness of *o* should not be confused with *believing* that *o* is red. To believe that *o* is red, one must predicate redness of *o* and do something like endorse the predication. In predicating redness of *o*, an agent merely represents *o* as red without committing herself to *o*’s redness. As we’ll see, this is a crucial difference between Soames’s theory and that of Hanks discussed below.

An agent predicating redness of *o* is a cognitive act token. Soames claims that the proposition that *o* is red is the act *type* of an agent predicating redness of *o*. Other more

⁴² See King, Soames, and Speaks (2014: ch. 10) and King (forthcoming) for details.

⁴³ Soames (2010a: 103).

complex propositions are identified with act types of agents performing *sequences* of primitive mental acts.⁴⁴ Soames holds that one entertains a proposition by performing an instance of the act type that it is.

As to “truth functional” propositions, Soames holds that the proposition that is the negation of the proposition *p* is the act type of representing each thing as having the property of *not being such that p*, the latter being a property that is had by either everything or nothing.⁴⁵ Similar remarks apply to conjunctive and disjunctive propositions. Quantified propositions are acts of predicating complex properties or propositional functions of other properties or propositional functions. The proposition that all *G*’s are *H*’s is the act of first applying a function to the property or propositional function *g*, yielding the property of *being true of all things of which g is true* and then predicating the latter of the property or propositional function *h*.

When an agent predicates redness of *o*, and so entertains the proposition that *o* is red, the agent *represents* *o* as being red. But this is a matter of *the agent* representing things being a certain way. *Propositions*, “act types” for Soames, are themselves supposed to represent things as being a certain way. But initially it may not be clear how that could be. In performing an act token of predicating redness of *o*, an agent represents *o* as being red precisely because she is predicating. In what sense could the *act type* be predicating redness of *o* and so represent *o* as red? Soames says that the act type represents *o* as being red in the derivative sense that for an agent to perform it is for the agent to represent *o* as being red.

On Soames’s account there are many more propositions than one might have expected. Consider the proposition that Plato was human. Call this proposition *P*. Suppose now I use the sentence “Plato was human” to perform *P* by understanding the sentence. I thereby use the name to pick out Plato, use the noun to pick out humanity, and use “was human” to predicate humanity of Plato. Using the sentence in this way is for Soames a representational cognitive act itself and so counts as a proposition *P** that is distinct from *P* (since entertaining *P* does not require e.g. using the name to pick out Plato, but entertaining *P** does). Entertaining *P** is a way of entertaining *P*, but not vice versa. Similarly, in understanding each of the following, I entertain the proposition that Hempel was a philosopher, which we’ll call *H*:

2a. Carl Hempel was a philosopher.

2b. Peter Hempel was a philosopher.

But in understanding 2a I also entertain a proposition *H_c* ($\neq H$) whose grasp requires me to identify Hempel by cognizing him using the name “Carl,” which proposition I do not entertain in understanding 2b. And in understanding 2b I also entertain a proposition *H_p* ($\neq H$ and $\neq H_c$) whose grasp requires me to cognize Hempel using the name

⁴⁴ Soames needs a number of primitive mental acts beyond predication. See Soames (2010a: 115, 122).

⁴⁵ Soames (forthcoming).

“Peter,” which proposition I do not entertain in understanding 2a. So in understanding both 2a and 2b I entertain three propositions, on Soames’s view: H , H_c , and H_p .⁴⁶ Presumably, this also means that when an English and Filipino speaker understand the sentences “Carl Hempel is a philosopher” and “Carl Hempel ay isang pilosopo,” respectively, though there is a common proposition they entertain (H above), each also entertains another proposition that the other does not (one involves using “is a philosopher” to predicate being a philosopher of Hempel, the other involves using “ay isang pilosopo” instead).

Jeff Speaks (King, Soames, and Speaks 2014) defends the view that propositions are properties. The proposition expressed by “Amelia talks,” for example, is the property *being such that Amelia instantiates talking*. Speaks holds that this property is expressed by the sentence in virtue of the fact that the syntactic concatenation in the sentence contributes the relation *___ is such that ___ instantiates ___* to the proposition expressed (as Speaks notes, this is his version of King’s notion of interpreting syntax), and the contents of “Amelia” and “talks” (Amelia and the property of talking, respectively) fill the second and third slots in this relation (in that order), yielding the property *being such that Amelia instantiates talking*. Truth for propositions is just instantiation: a proposition is true iff it is instantiated. Further, if a proposition is instantiated, everything instantiates it. If the actual world is such that Amelia instantiates talking, then so are you, Paris, and everything else. Further, a proposition P is true at a world w iff if w were actual, P would be instantiated.⁴⁷

Since we want there to be false and even necessarily false propositions, Speaks’s account requires uninstantiated properties and even properties that couldn’t be instantiated. Speaks notes that this is a cost of his account but the price doesn’t seem terribly high. Obviously, it does mean, though, that no one whose theory of properties eschews uninstantiated properties can adopt Speaks’s account.

Turning to the propositional attitudes, according to Speaks, to believe that Amelia talks is to bear an attitude toward the property *being such that Amelia talks*. In particular, he claims it is to believe something is such that Amelia talks. But putting things this way obscures the fact that for Speaks belief is a relation between an individual and a property (what appears to be said to be believed here is that something is such that Amelia talks, and that doesn’t sound like a property). Perhaps it is better to make explicit that belief is a relation to a property for Speaks and say that to believe that Amelia talks is to bear the believes-instantiated relation to the property *being such that Amelia talks*. Speaks holds that one may believe-instantiated a property because there is some particular thing that one believes instantiates the property. He suggests that one might believe-instantiated the property that Amelia talks because one believes that the world instantiates that property. He suggests that this fits well with the view that in believing, one takes the world to be a certain way.

⁴⁶ Soames (forthcoming).

⁴⁷ Actually, Speaks expresses doubt as to whether this is what he wants to say ultimately. See King, Soames, and Speaks (2014: ch. 5 fn. 8).

There are other cognitive relations we bear to properties that are not propositions on Speaks's view. Here he wants to give an account of "first person" mental states that others have called belief, desire, etc. *de se*. Agents can bear the relation of self-attribution to properties like *being on fire*. This gives us "first personal" mental states of the sort discussed in the literature on *de se* belief. But note that Speaks's account here is not one of *de se* belief because the properties self-attributed are not propositions and self-attribution of properties is not belief for Speaks. On his account, he could believe Jeff Speaks is on fire, which for Speaks amounts to his believing-instantiated the property *being such that JS is on fire*. Or he could take himself to be on fire, which for him amounts to self-attributing the property *being on fire*. The former is Speaks standing in a relation to a proposition; the latter is not.

As to the semantics of verbs of propositional attitude, Speaks takes "believes" to express the following two-place relation between a person and a property: ____ takes to be instantiated the property ____ (similar remarks apply to other verbs of attitude). It is important to see that for Speaks belief ascriptions, for example, do not by means of their semantics ever assert that one self-attributes a property (similar remarks hold for desire ascriptions etc.). Thus for Speaks, both "I believe JS is on fire" and "I believe I am on fire" (uttered by Speaks) express the proposition that JS stands in the believes-instantiated relation to the property (proposition) *being such that JS is on fire*. Speaks suggests that perhaps the proposition that JS self-attributes being on fire can be pragmatically conveyed by his uttering "I believe that I am on fire." But Speaks is upfront about the fact that he has no explanation of the mechanism by means of which this proposition is pragmatically conveyed by uttering the sentence in question, while acknowledging that an account of the mechanism is required in order to sustain the suggested pragmatic approach.

A final point about Speaks's view. Though for him propositions *qua* properties have truth conditions ("instantiation conditions"), they do not represent anything and are not about anything. But surely our mental states with propositional content, such as believing San Clemente is beautiful, *do* represent and *are* about something? How can this be, however, on Speaks's view, if the things that are their contents don't have representational properties? Speaks holds that what makes my belief that San Clemente is beautiful about something is not the object of my belief (a proposition *qua* property) but the relation I stand in to it. Mental states like belief with representational properties can be factored into two parts: *a relation* that one bears to *the propositional content*. While most people have claimed that the latter infuses the mental state with its representational properties, Speaks claims it is (primarily) the former.⁴⁸

Peter Hanks (2015) defends a view with certain similarities to Soames's discussed above. Like Soames, Hanks follows King in endorsing the three novel claims from King (2007) and subsequent work mentioned above: he (i) rejects theories according to

⁴⁸ The qualification is here because, as Speaks notes, the representational properties I instantiate when I believe snow is white have to be different from those I instantiate when I believe grass is green, and this difference is attributable to the difference in the propositions believed.

which propositions have truth conditions and so represent the world as being a certain way by their very natures and independently of minds and languages as mysterious; (ii) holds that an adequate theory of propositions must *explain* how/why propositions have truth conditions and so represent the world as being a certain way; and (iii) holds that since the representational capacities of propositions cannot be something they have inherently and by their natures, their representational capacities must derive from and be explained by the representational capacities of thinking agents.

In simple cases of judging and asserting, Hanks claims we perform actions of predicating properties of objects. In silently judging that Ted Cruz is a demagogue, one thinks of Cruz (refers to him in thought), thinks of the property of being a demagogue (expresses the property in thought), and predicates the property of Cruz. In asserting that Cruz is a demagogue, one likewise refers to him, expresses the property, and predicates it of him, this time by linguistic means. According to Hanks, in predicating a property of an object, one characterizes the object as being a certain way and so does something that is true or false. That is, Hanks claims these token acts of predicating are true or false. Unlike Soames's notion of predication, Hanks's notion is inherently assertive and committing. There is not some neutral thing you do in predicating and then add endorsement or commitment. Predicating *is* committing to the object possessing the property in question on Hanks's view. The proposition that Ted Cruz is a demagogue is the act *type* of predicating *being a demagogue* of Cruz. Hanks holds that this act type would exist even if it had no tokens. In general, he holds that the act types that are propositions exist whether they have tokens or not.

As we've seen, the proposition that Cruz is a demagogue is the act type of referring to Cruz, expressing the property of being a demagogue and predicating it of Cruz. What Hanks calls the *interrogative proposition* expressible by the question "Is Cruz a demagogue?" is the act type of referring to Cruz, expressing the property of being a demagogue and combining Cruz and the property in an interrogative way (rather than *predicating* the property of Cruz). Similarly, the *imperative proposition* expressed by the command "Cruz, be a demagogue!" is the act type of referring to Cruz, expressing the property of being a demagogue and combining object and property in an imperative manner. It should be clear that Hanks rejects that there is some common underlying content running through the (assertive) proposition, imperative proposition, and interrogative proposition regarding Cruz being a demagogue. There are just three different propositions here with their "forces" (assertive, interrogative, imperative) built in.

Hanks's explanation of why propositions as act types of the sort described have truth conditions comes in two steps. First, he argues that token acts of predicating properties of objects have truth conditions. He recognizes that some may doubt whether act tokens are the kinds of things that can be true and false. His main argument that they can be is that we have adverbial modifiers "truly" and "falsely" and that these attribute properties to action tokens. So just as "quickly" in "Obama quickly stated that Clinton is eloquent" attributes a property to Obama's action, so Hanks claims that "truly" in "Obama truly stated that Clinton is eloquent" attributes the property of being

true to an action of Obama. Hanks then tries to argue that the action *type* of predicating being a demagogue of Cruz inherits the property of having truth conditions from its instances. Note that here it must be actual and possible instances that truth conditions are inherited from, since Hanks wants propositions that have never been and never will be tokened to have truth conditions. The argument here is complex since Hanks notes that types inherit certain kinds of property from tokens and not others. Hence, Hanks seeks to explain why the having of truth conditions is the kind of property a type inherits from its (actual and possible) tokens.

Just as on Soames's theory, on Hanks's there are many more propositions than one might have thought. First, there is the proposition that is the act type consisting of referring to Clinton using "Clinton" and predicating eloquence of her. It is true iff Clinton is eloquent. Then there is a distinct proposition with those same truth conditions that is the act of referring to Clinton in any way whatsoever and predicating eloquence of her. Next, there is another proposition with these same truth conditions consisting of the act of referring to Clinton using "Clinton" while thinking of her as Obama's former secretary of state and predicating eloquence of her. In addition, there is another proposition with the same truth conditions consisting of referring to Clinton using "Clinton" while thinking of her as a former first lady and predicating eloquence of her. And so on. Again, all these propositions are true iff Clinton is eloquent. There is even a proposition consisting of referring to Clinton using "Clinton" while drawing a round square and predicating eloquence of her.⁴⁹

Above we saw that for Hanks, predication endows propositions with an inherent element of judgment or assertion. Predicating a property of an object *commits* the predicator to the object having that property. For Hanks, she did something false if it doesn't. But this raises a problem for him. In disjunctive, negated, conditional propositions and others, embedded propositions do not have assertive force. They aren't asserted. But if these embedded propositions inherently have an assertoric or judgmental element, how can this be? To address this Hanks claims that "or," "if," and "not" create *cancellation contexts*. Since "or" for example creates a cancellation context on Hanks's view, in uttering the following sentence I fail to assert that Cruz is a demagogue:

Cruz is a demagogue or Clinton is eloquent.

Each disjunct expresses a proposition that includes as a sub-act the act of predicating a property of the object, according to Hanks. But the predication is canceled. Hence, the asserting/committing element in the proposition is canceled. Hanks must walk a very fine line here. The inherently assertive committal aspect of predication is what explains both that propositions are true and false and that unembedded propositions make assertions.⁵⁰ To explain why some embedded propositions *don't* make assertions,

⁴⁹ See Hanks (2015: 77–8 and 78 fn. 10).

⁵⁰ Hanks (2015: 22).

Hanks appeals to the notion of the predication being canceled in virtue of occurring in a cancellation context. But of course Hanks will want to say that the embedded proposition still has truth conditions and is true or false. Somehow in canceling the predication, though assertive force is canceled the truth conditions and truth-value aren't, even though predication is responsible for both.

Above it was mentioned that King (2007) and subsequent work endorsed three novel claims about propositions. Two of these are: (i) theories according to which propositions have truth conditions and so represent the world as being a certain way by their very natures and independently of minds and languages are mysterious and to be rejected; (ii) an adequate theory of propositions must *explain* how/why propositions have truth conditions and so represent the world as being a certain way.⁵¹ It was also noted that Soames and Hanks follow King in this respect. In a sense, Speaks does too insofar as he feels the need to explain the "aboutness" or intentionality of some mental states.⁵² So most recent literature has taken on board King's two claims above. Philosophy being what it is, however, there have been recent theorists who want to return to something like the classical theories of Frege and Russell, where propositions were thought to be mind- and language-independent abstract objects that by their very natures had truth conditions and so represent the world as being a certain way. On such a view, both of King's claims are rejected. Propositions are precisely held to have truth conditions by their natures and independently of minds and languages; and it is denied that there can be any explanation of how/why propositions have truth conditions. Trenton Merricks (2015) defends such a view. According to Merricks, propositions are abstract, necessary existents that essentially represent the world as being a certain way.⁵³ He also holds that the fact that propositions essentially represent things as being a certain way is primitive. That is, this fact has no explanation.⁵⁴ Further, Merricks holds that propositions are *simple*: they have no constituents.⁵⁵ On this last point at least Merricks departs from the classical views of Frege and Russell on which propositions do have constituents.

Lorraine Juliano Keller (2014) expresses sympathy for a similar view that she calls *propositional primitivism*. On this view, propositions are fine-grained *sui generis* entities. They are not reducible to nor can they be explained by entities in another ontological category. Her primitivist thinks that not much of anything can be said about the inner nature of propositions except that they are abstract, mind- and language-independent entities with no constituents or structure. While King, Soames, Hanks, and perhaps Speaks see in such a view mystery or worse, primitivists like Keller and Merricks see elegance and simplicity.

⁵¹ King actually made this point in earlier work. See n. 38.

⁵² Recall that on Speaks's view, the aboutness of propositions doesn't have to be explained, because they aren't about anything.

⁵³ Merricks (2015: 191–4).

⁵⁴ Merricks (2015: 195).

⁵⁵ Merricks (2015: 207).

Finally, we close with some brief remarks on how the philosophers we have discussed think of the truth of propositions and what makes them true. Frege held that the truth as applied to thoughts was *sui generis* and indefinable.⁵⁶ Hence nothing substantive can be said about it. Russell (1903) struggles with the question of how true propositions differ from false ones. He had the idea that there was a *logical sense of assertion* in which only true propositions are asserted but he was unable to say anything substantial about it and in the end left the nature of truth to “the logicians.”⁵⁷

Once propositions were construed as sets of possible worlds, there was a very simple account of the truth: a proposition P is true at world w iff $w \in P$. Obviously, though, this doesn’t seem very deep. But it does represent a change in the sense that propositions came to be seen as true and false *relative to possible worlds*.

As to the algebraic approaches, recall that Bealer (1993a; 1998) has a set of extensionalization functions that includes a distinguished extensionalization function H that assigns properties, n -place relations, and propositions their actual extensions. Where $\text{pred}_s(F, o)$ is the proposition that singularly predicates property F of individual o , $H(\text{pred}_s(F, o)) = T$ iff $o \in H(F)$. But $o \in H(F)$ just in case o has property F . Thus Bealer’s propositions are made true by things being a certain way (objects possessing properties, etc.). Structured proposition theorists in the 1980s like Soames (1987) gave definitions of *truth at a world* for propositions. Where P^* is an n -place relation and o_1, \dots, o_n are n individuals, we get clauses in that definition such as

A proposition $\langle\langle o_1, \dots, o_n \rangle P^*\rangle$ is true relative to a circumstance E iff the extension of P^* in E contains $\langle o_1, \dots, o_n \rangle$.

Looking at clauses like this, it is natural to think that a structured proposition represents the world as being a certain way by having its propositional constituents configured a certain way in the proposition. Since these constituents are things like objects, properties, and relations, we can say that by configuring objects o, o' and the two-place relation R a certain way in the proposition, the proposition represents the world as being one in which oRo' .

Finally, despite the differences between their views, all of King, Soames, Speaks, Hanks, Merricks, and Keller take propositions to have truth conditions and so to be true or false relative to possible worlds.⁵⁸ On all these views, propositions impose conditions on a world that must be met for them to be true relative to the world. All these authors except Speaks take propositions to represent the world as being a certain way, and claim that propositions are true at worlds when they are the way the proposition represents them as being.

⁵⁶ Frege (1918: 327).

⁵⁷ Russell (1938 [1903]: section 52).

⁵⁸ Merricks (2015) thinks some propositions can vary in truth-value over time. See (2015: 112).

REFERENCES

- Bealer, G. (1993a). "A Solution to Frege's Puzzle," *Philosophical Perspectives* 7: 17–60.
- Bealer, G. (1993b). "Universals," *Journal of Philosophy* 90: 5–32.
- Bealer, G. (1998). "Propositions," *Mind* 107: 1–32.
- Beaney, M. (1997). *The Frege Reader*. Oxford: Blackwell.
- Collins, J. (2007). "Syntax, More or Less," *Mind* 166: 805–50.
- Frege, G. (1964). *The Basic Laws of Arithmetic*, ed. and trans. M. Furth. Berkeley: University of California Press.
- Frege, G. (1891a). "Function and Concept," reprinted in Beaney (1997), 84–129.
- Frege, G. (1891b). "Comments on Sinn and Bedeutung," reprinted in Beaney (1997), 172–80.
- Frege, G. (1892a). "On Sinn and Bedeutung," reprinted in Beaney (1997), 151–71.
- Frege, G. (1892b). "On Concept and Object," reprinted in Beaney (1997), 181–93.
- Frege, G. (1918). "Thought," reprinted in Beaney (1997), 325–45.
- Frege, G. (1919). "Notes for Ludwig Darmstaedter," reprinted in Beaney (1997), 362–7.
- Hanks, P. (2015). *Propositional Content*. Oxford: Oxford University Press.
- Kaplan, D. (1977). "Demonstratives," reprinted in J. Almog, J. Perry, and H. Wettstein (eds.) (1989), *Themes from Kaplan*. New York: Oxford University Press, 481–563.
- Keller, L. J. (2014). "The Metaphysics of Propositional Constituency," *Canadian Journal of Philosophy* 43: 655–78.
- King, J. C. (1995). "Structured Propositions and Complex Predicates," *Noûs* 29: 516–35.
- King, J. C. (2007). *The Nature and Structure of Content*. Oxford: Oxford University Press.
- King, J. C. (2009). "Questions of Unity," *Proceedings of the Aristotelian Society* 109(3): 257–77.
- King, J. C. (2013). "On Fineness of Grain," *Philosophical Studies* 163: 763–81.
- King, J. C. (2013). "Propositional Unity: What's the Problem, Who Has it and Who Solves it?," *Philosophical Studies* 165: 71–93.
- King, J. C., S. Soames, and J. Speaks. (2014). *New Thinking About Propositions*. Oxford: Oxford University Press.
- King, J. C. (forthcoming). "On Propositions and Fineness of Grain (Again!)," *Synthese* special issue, ed. M. G. Carpintero and B. Jespersen.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, D. (1970). "General Semantics," *Synthese* 22: 18–67.
- Menzel, C. (1993). "The Proper Treatment of Predication in Fine-Grained Intensional Logic," *Philosophical Perspectives* 7: 61–88.
- Merricks, T. (2015). *Propositions*. Oxford: Oxford University Press.
- Moltmann, F. (2014). "Propositions, Attitudinal Objects, and the Distinction between Actions and Products," *Canadian Journal of Philosophy*, supplementary volume on propositions, ed. G. Rattan and D. Hunter, 43.5–6: 679–701.
- Montague, R. (1968). "The Proper Treatment of Quantification in Ordinary English," reprinted in R. H. Thomason (ed.), *Formal Philosophy: Selected Papers of Richard Montague* (1974). New Haven: Yale University Press, 247–70.
- Russell, B. (1938) [1903]. *The Principles of Mathematics*. New York: W. W. Norton.
- Russell, B. (1994) [1910]. "On the Nature of Truth and Falsehood," in B. Russell, *Philosophical Essays* (reprint edn). New York: Routledge.
- Richard, M. (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. New York: Cambridge University Press.

- Salmon, N. (1986). *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Schiffer, S. (2003). *The Things We Mean*. Oxford: Oxford University Press.
- Soames, S. (1987). "Direct Reference, Propositional Attitudes and Semantic Content," *Philosophical Topics* 15: 47–87. Reprinted in N. Salmon and S. Soames (eds.) (1988), *Propositions and Attitudes*. New York: Oxford University Press, 197–239.
- Soames, S. (2010). *What is Meaning?* Princeton: Princeton University Press.
- Soames, S. (2015). *Rethinking Language, Mind, and Meaning: The Hempel Lectures*. Princeton: Princeton University Press.
- Soames, S. (forthcoming). "Propositions as Cognitive Acts," *Synthese* special issue, ed. M. G. Carpintero and B. Jespersen.
- Stalnaker, R. (1987). *Inquiry*. Cambridge, MA: MIT Press.
- Stalnaker, R. (1999). *Context and Content*. New York: Oxford University Press.
- Stalnaker, R. (2003). *Ways a World Might Be*. New York: Oxford University Press.
- Zalta, E. (1988). *Intensional Logic and the Metaphysics of Intentionality*. Cambridge, MA: MIT Press.

CHAPTER 13

TRUTHMAKERS

ROSS P. CAMERON

13.1 INTRODUCTION

TRUTHMAKER theory says that there is an intimate link between truth and ontology: that is, between what is the case and what there is.¹ According to truthmaker theory, for a proposition to be true requires there to be some things that *make* it true. The truthmakers are the ontological ground of the truth; their existence explains *why* the proposition in question is true.

For example, it is true that there were dinosaurs. Why? The truthmaker theorist says we can't just take this historical fact as brute—we must provide an ontological explanation for its truth: there must be some things that *make* it the case that there were dinosaurs, whose existence explains why the historical facts are as they are in this respect. Truth doesn't come for free, it must be grounded in ontology—that is the truthmaker theorist's thought. This chapter will explore some of the issues surrounding truthmaker theory: what principles govern truthmaking, why we should believe the theory, and connections between truthmaker theory and other debates concerning truth and ontology.

What is it for some things to make a proposition true? It's not the same as a person making an artifact, such as a house. The connection between truthmakers and proposition is not a causal one. Why not? Well, on many accounts of propositions they're abstract entities, and simply not the kind of thing that can enter into causal relationships.² Also, in general, causal connections are contingent connections, in the sense that the entities related by cause and effect might both have existed but not been so related—my kicking of the ball caused the breaking of the window, but both might have existed and not been so related, had you caught the ball but something

¹ Thanks to Elizabeth Barnes, Michael Glanzberg, John Heil, Alex Skiles, and Robbie Williams for helpful comments.

² For a discussion of the metaphysics of propositions, see King (ch. 12 in this volume).

else broken the window; but the truthmaking relation is usually taken as a relation of necessitation—that if some things make a proposition *p* true then it's simply *impossible* for them to exist and *p* be false. The mere existence of the truthmakers is sufficient for the truth of the proposition in question. We'll look below at whether or not this claim—truthmaker necessitarianism—is true; but if it is then that is another reason to think that the truthmaking relation is not a causal one. Given that the causal facts could have differed, causation calls out for grounding, rather than providing the grounding.

If the truthmaking relation is one of necessitation, then perhaps that's simply what truthmaking *is*: perhaps all it is for the *Xs* to make *p* true is for the mere existence of the *Xs* to suffice for the truth of *p*. That would be a neat account if true, for then we could analyze truthmaking using the relatively familiar notion of necessity. Unfortunately, it doesn't seem to be a sufficient condition for the *Xs* to make *p* true that their existence necessitate the truth of *p*. For one, this seems hopelessly wrong in the case of necessary truths. *Every* entity is such that its existence necessitates the truth of every necessary truth, but we might well balk at saying that every entity makes every necessary truth true. The sufficiency claim also generates absurd results if certain essentialist theses are true; suppose, for example, that the essentiality of origin is true, and I couldn't have been born of different parents—in that case, my existence necessitates their existence, but it's not the case that I am the truthmaker for the fact that they exist: truthmaking is meant to be explanatory, but my existence does not explain why my parents exist. Most importantly, necessitation is blind to what truthmaking should *not* be blind to: the direction of explanation. Two theorists might agree for instance that there is some thing—the state of affairs of electron *e* being negatively charged³—whose existence necessitates the truth of $\langle e \text{ is negatively charged} \rangle$.⁴ But this is silent on which comes first: the truth of the proposition or the existence of the state of affairs.⁵ One theorist may hold that the state of affairs exists *because* of how things are, while the other thinks that things are as they are because of this state of affairs existing. Truthmaking demands the latter: what's true is true *because* of what exists, so the existence of the truthmakers had better come prior to the truth of the propositions, rather than their existence being dependent on the truth of the propositions in question. We need an account of truthmaking that will let us talk about metaphysical priority/dependence/explanation, and so on if we want to capture the truthmaker theorist's thought that what's true is *grounded in* or *depends on* what there is. Modal notions are not up to the task.

³ If truthmaker necessitarianism is true then, assuming that ordinary objects like you and I, tables and chairs, etc. have many of their properties accidentally (i.e. they could exist but have different properties), ordinary objects will not be sufficient truthmakers for most of the true predications concerning them: the truthmaker theorist needs to invoke more exotic objects like states of affairs—an entity that binds an object and its properties together—or tropes—properties of particular individuals that couldn't belong to anything else. See Armstrong (1997; 2004) for discussion of the options.

⁴ I use " $\langle p \rangle$ " to refer to the proposition that *p* is the case.

⁵ Cf. Nolan (2008).

Since truthmaking appears to be neither causation nor necessitation, we might be tempted to take it as primitive, but I think that would be a mistake. As well as the phenomenon of some things making a proposition true, there is also the phenomenon of one truth being true in virtue of some others. For example, that it is wrong to do X is, some think, true in virtue of the fact that there is a possible action Y that could be performed that would have better consequences than the consequences X would have if performed. It would be uneconomical if we had to have two primitive notions here, so we should try and define one of *makes true* and *in virtue of* in terms of the other.

I think the prospects of defining the *in virtue of* relation in terms of the truthmaking relation are dim, however. How would it go? Certainly not “p is true in virtue of q iff q makes p true.” For one thing, any account that invokes propositions as truthmakers either has to deny truthmaker necessitarianism, deny that propositions are necessary existents or deny that contingent truths have truthmakers. Secondly, even if propositions are contingent existents, they don’t need to be *true* to exist. Suppose <X is wrong> is, as the consequentialist says, true in virtue of <X would have bad consequences>; the definition tells us that the latter proposition makes it true that X is wrong. But that proposition can exist and be false (there are worlds where doing X would have good consequences), so now we’re committed to thinking that in a world where X would have good consequences, X is still wrong, which is exactly the opposite of what the consequentialist wants. What about “p is true in virtue of q iff the truth of q makes p true”? That solves the above problem; but what is this entity *the truth of q*? Even if we believe in propositions, we might be hesitant to believe in some thing that is the particular truth of a particular proposition. What about “p is true in virtue of q iff the state of affairs that q makes p true”? That might work if there was a state of affairs that p for every true proposition p; but on the most plausible accounts of states of affairs there is not, and in any case we shouldn’t build it into our conception of truthmaking that this is the case. “p is true in virtue of q iff, necessarily, whatever makes q true makes p true”? No: it’s no part of the *definition* of truthmaking that every truth has a truthmaker, and we should allow for the possibilities both that one proposition is true in virtue of another even though neither have truthmakers, and that two propositions lack truthmakers but where one is *not* true in virtue of the other. And if every truth *does* have a truthmaker, the definition will entail the wrong result that <there is something> is true in virtue of any true negative existential, such as <there are no unicorns>, since whatever makes the negative existential true necessarily makes it true that there is something, since *any things* that make anything true necessarily makes it true that something exists, since they make it true that *they* themselves exist.

But truthmaking *can* be defined in terms of *in virtue of*. Truthmaker theory says that what is true is grounded in what there is: as I will understand it, this is the claim that the totality of truths are ultimately true in virtue of just those truths that are concerned solely with ontology—that is, that any truth at all is ultimately true in virtue of some truths concerning what there is.

13.2 TRUTHMAKING AND *IN VIRTUE OF*

I will take as primitive, then, the *in virtue of* relation that holds between true propositions.^{6,7} Truthmaking will be defined in terms of this relation below. The *in virtue of* relation is irreflexive, asymmetric, and transitive. The converse of this relation is *grounding*: $\langle q \rangle$ grounds $\langle p \rangle$ just in case $\langle p \rangle$ is true in virtue of $\langle q \rangle$. Let us write “ $q \Rightarrow p$ ” for “ $\langle p \rangle$ is true in virtue of $\langle q \rangle$,” or “ $\langle q \rangle$ grounds $\langle p \rangle$.” The three conditions on the *in virtue of* relation can then be written as

Irreflexivity: $\neg \exists x(x \Rightarrow x)$

Nothing is its own ground. You don’t provide an explanation for the truth of $\langle p \rangle$ by citing the truth of $\langle p \rangle$. If x grounds y , then, x and y are distinct true propositions.

Asymmetry: $\forall x \forall y(x \Rightarrow y \rightarrow \neg(y \Rightarrow x))$

There cannot be explanatory circles; at least, not when it is metaphysical explanation we are concerned with. Grounding takes us from the less fundamental to the more fundamental: we can’t keep going and end up back where we started. If x is true in virtue of y then the truth of y is prior to the truth of x , and so y can’t also be true in virtue of x .

Transitivity: $\forall x \forall y \forall z((x \Rightarrow y \ \& \ y \Rightarrow z) \rightarrow x \Rightarrow z)$

⁶ I assume that *in virtue of* is one-one: that it always takes you from a single proposition that is derivative to a single proposition that grounds it. Nothing really hangs on this; it just makes the presentation easier—as we will see, the resulting definition of truthmaking still allows that a proposition may be made true by many things rather than by one single thing. If you think that a proposition can be true in virtue of *some* propositions, or even that some propositions can be true in virtue of some propositions, everything I say can be modified to allow for that.

⁷ I have heard it objected to this approach that this requires us to take propositions as fundamental entities. The objection is misplaced: taking propositions to be the relata of the *in virtue of* relation is entirely neutral on the question as to whether propositions themselves are fundamental or derivative. The approach entails that any chain of dependence ends (if it ends at all) at some proposition, but it simply does not follow that that proposition is a fundamental existent: all that follows from an *in virtue of* chain ending in a proposition Q is that *what Q says* to be the case is fundamentally the case; but *that Q exists* needn’t be fundamentally the case! My view is consistent, for example, with taking every proposition to be derivative, with the fact that it exists being true in virtue of the fact that it is possible for someone to entertain the content of that proposition. So now suppose P holds in virtue of Q , but Q does not hold in virtue of anything; then *what Q says* to be the case is fundamentally the case, but *that Q exists* obtains in virtue of the proposition (call it R) that someone could entertain Q ’s content. Of course, now we have invoked another proposition, R ; but that R exists will likewise not be fundamental but true in virtue of the proposition that someone could entertain *its* content. And so on, and so on. This generates an infinite chain of *in virtue of* relations, since each instance of the relation invokes a new proposition whose existence must be grounded in a new instance of the relation. But this is not a vicious regress, because at no point does the success of an instance of *in virtue of* depend on the success of the new instance generated; it matters not to P ’s being grounded in Q that Q is a derivative entity.

A proposition x is true in virtue of any proposition z that grounds a proposition y that grounds x . $\langle p \rangle$ can be true in virtue of some truth, $\langle q \rangle$, which is *more* fundamental than $\langle p \rangle$ but not absolutely fundamental; i.e. $q \Rightarrow p$, but there is some true proposition, $\langle r \rangle$, such that $r \Rightarrow q$. In that case $\langle q \rangle$ does not provide the ultimate grounds for $\langle p \rangle$; the truth of $\langle q \rangle$ does not provide the most general explanation for the truth of $\langle p \rangle$, since the truth of $\langle q \rangle$ is itself something that needs to be explained by the truth of $\langle r \rangle$. Since $\langle r \rangle$ explains why $\langle q \rangle$ is true and $\langle q \rangle$ explains why $\langle p \rangle$ is true, there is a perfectly good sense in which $\langle r \rangle$ explains why $\langle p \rangle$ is true—so we say that $\langle p \rangle$ is true in virtue of $\langle r \rangle$. If $\langle r \rangle$ is itself brute then $\langle r \rangle$ provides the ultimate grounds for $\langle p \rangle$.

Definition of “brute”: A brute truth is a true proposition that has no ground—that is not true in virtue of anything. $\langle p \rangle$ is *brute* iff_{df} p is true and $\neg \exists x(x \Rightarrow p)$. A proposition is derivative iff it is true and not brute.

Definition of “ultimate ground”: $\langle p \rangle$ is an *ultimate ground* for $\langle q \rangle$ iff_{df} $p \Rightarrow q$ and $\langle p \rangle$ is brute.

The transitivity of *in virtue of* is stipulative. There is an interesting non-transitive relation that holds between $\langle p \rangle$ and $\langle q \rangle$ and between $\langle q \rangle$ and $\langle r \rangle$ but doesn't hold between $\langle p \rangle$ and $\langle r \rangle$: call this *direct grounding*. The reason I am taking grounding and not direct grounding as primitive is that the latter can be defined in terms of the former but not vice versa. The definition of the non-transitive notion in terms of the transitive one is as follows.

Definition of “direct grounding”: $\langle p \rangle$ *directly grounds* $\langle q \rangle$ iff (i) $\langle p \rangle$ grounds $\langle q \rangle$, and (ii) there is no $\langle r \rangle$ such that (a) $\langle p \rangle$ grounds $\langle r \rangle$, and (b) $\langle r \rangle$ grounds $\langle q \rangle$.⁸

And “directly in virtue of” is, of course, defined as the converse of direct grounding. But grounding doesn't appear to be definable in terms of direct grounding, hence the reason for taking grounding as primitive. We might try the following: $\langle p \rangle$ grounds $\langle q \rangle$ iff $\langle p \rangle$ directly grounds $\langle q \rangle$ or $\exists x(\langle p \rangle$ directly grounds x and x directly grounds $\langle q \rangle$) or $\exists x \exists y(\langle p \rangle$ directly grounds x and x directly grounds y and y directly grounds $\langle q \rangle$) or etc. But this assumes that for every case of grounding there is a non-dense chain of direct groundings linking the ground proposition to the grounding proposition, and this assumption looks wholly unjustified. Indeed, it looks like there can be grounding even when there is no direct grounding at all, making the prospects of defining grounding in terms of direct grounding hopeless. For what is to rule out there being a mapping from the positive real numbers to propositions, such that a proposition x grounds a

⁸ This definition is good if no proposition ever both directly *and* indirectly grounds some other proposition, but it will not allow for cases where there is both a direct grounding from $\langle p \rangle$ to $\langle q \rangle$ but where *in addition* $\langle p \rangle$ also grounds $\langle q \rangle$ via grounding some $\langle r \rangle$ which grounds $\langle q \rangle$. I'm not convinced we *should* allow for such cases, but perhaps they shouldn't be ruled out by definition. If so, we should not talk about whether $\langle p \rangle$ directly grounds $\langle q \rangle$ *simpliciter*, but rather whether a particular *instance* of grounding is direct or indirect: a particular instance of $\langle p \rangle$ grounding $\langle q \rangle$ will be direct iff *that instance* of grounding isn't mediated via some $\langle r \rangle$ which grounds $\langle q \rangle$ and is grounded by $\langle p \rangle$. Thanks to Alex Skiles for pressing me on this.

proposition y just in case the real number mapped on to x is less than the real number mapped on to y ? In such a situation, there are no instances of direct grounding: whenever x grounds y , there will be some proposition, z , that grounds y and is grounded by x . Since such situations look conceptually possible, we should not attempt to define the transitive notion of grounding in terms of the non-transitive notion of direct grounding; rather we should take the transitive notion as primitive and define direct grounding as I do above.

Call the set containing all and only the brute propositions Π . If Π is empty then every proposition is derivative, in which case no proposition has an ultimate ground.

Consider also the set—call it Σ —of propositions whose entire content is that some thing, or some things, exist(s): call these propositions *pure existence claims*. (Pure existence claims will be expressible by sentences of the form “ a exists” or “the X s exist,” where “ a ” is a rigid designator and “the X s” a rigid plurally referring expression (i.e. it plurally refers in every possible world to the things that are actually the X s, if they exist, and it fails to refer if any of the actual X s fail to exist.)

We can now define truthmaking as follows:

Truthmaking: A proposition $\langle p \rangle$ is made true by X , or the X s, just in case *either* (i) $\langle p \rangle \in \Pi$ & $\langle p \rangle \in \Sigma$ & $\langle p \rangle$ says that X (or the X s) exist(s) *or* (ii) $\exists x(x \Rightarrow p \ \& \ x \in \Pi \ \& \ x \in \Sigma \ \& \ x$ says that X (or the X s) exist(s)).

That is: a proposition is made true by some things, the X s, if and only if it is the brutally true pure existence claim that the X s exist or it is true in virtue of the brutally true pure existence claim that the X s exist.

Truthmaker theory, as I’ll understand it, involves two claims. The first is that the only brute truths are pure existence claims; i.e. that the brute truths are all propositions whose sole content concerns the existence of some thing(s). The second claim is that every proposition is either brute or has an ultimate grounding. I will call these claims “Priority of Existence” and “Universal Grounding” respectively.

Priority of Existence: $\Pi \subseteq \Sigma$. The brute truths are a subset of the pure existence claims.

Universal Grounding:⁹ $\forall p(\neg \exists x(x \Rightarrow p) \vee \exists y[(y \Rightarrow p) \ \& \ \neg \exists z(z \Rightarrow y)])$. Every true proposition is either brute or is true in virtue of some brute proposition.

Universal Grounding tells us that there is a class of brute truths that provide the ultimate grounding for every other truth, and Priority of Existence tells us that this class only includes the pure existence claims. Combined, then, they tell us that every true proposition is either a pure existence claim or is true in virtue of a pure existence claim; by the above definition of truthmaking, this is the characteristic claim of truthmaker theory: the claim that every truth is made true by some thing(s).

⁹ The universal quantifier here ranges over true propositions.

If *every* proposition in Σ is made true by the thing or things it says exist, then $\Sigma \subseteq \Pi$, from which it follows, given Priority of Existence, that $\Pi = \Sigma$. That is, not only will the brute truths only include the pure existence claims, they will include *all and only* the pure existence claims—the set of brute truths and the set of pure existence claims will be co-extensive (hence, identical). But I don't think we should accept this; I think that there are pure existence claims that are true in virtue of other pure existence claims. In Cameron (2008a) I argue that propositions proclaiming the existence of sets, or complex objects, may be true solely in virtue of the propositions proclaiming the existence of the members and simple parts, respectively. The advantage of allowing for such cases is that it allows us to make sense of some entities being an “ontological free lunch”: the As are no ontological addition to the Bs, and hence one incurs no further ontological commitments in accepting the As over accepting the Bs, if $\langle \text{The As exist} \rangle$ is true in virtue of $\langle \text{The Bs exist} \rangle$: so sets and complex objects would be, in this sense, no addition in being over their members/parts.

13.3 TRUTHMAKER NECESSITARIANISM

As we saw above, it is common for truthmaker theorists to hold the doctrine known as truthmaker necessitarianism: that the existence of the truthmaker(s) for a proposition $\langle p \rangle$ couldn't exist and $\langle p \rangle$ fail to be true. But why? David Armstrong offers us a proof of necessitarianism. He says,¹⁰

If it is said that the truthmaker for a truth could have failed to make the truth true, then we will surely think that the alleged truthmaker was insufficient by itself and requires to be supplemented in some way. A contingently sufficient truthmaker will [make the truth] true only in circumstances that obtain in this world. But then these circumstances, whatever they are, must be added to give the full truthmaker.

His thought, I take it, is this. Suppose that the Xs make $\langle p \rangle$ true but don't necessitate the truth of $\langle p \rangle$. In that case there are some possible worlds in which the Xs exist and $\langle p \rangle$ is false: some set of circumstances in which the existence of the Xs does not suffice for the truth of $\langle p \rangle$. But then isn't it overwhelmingly intuitive that the truthmaker for $\langle p \rangle$ is not simply the Xs but rather them together with whatever makes it true that those circumstances do *not* in fact obtain?

Unfortunately, I don't think this argument works. Ask yourself this: why isn't it simply the Xs themselves that make it true that those circumstances do not in fact obtain? Why think that something has to be *added* to ensure that the actual circumstances are ones in which the Xs will make p true—why can't the Xs ensure this on their own?

¹⁰ Armstrong (1997: 116). He gives the same argument in Armstrong (2004: 6–7).

Armstrong's only reason, it seems, is that the Xs can't ensure that the circumstances are right for them to make *p* true because, *ex hypothesi*, they could exist in circumstances that aren't right. But one can only conclude from this that they doesn't actually make it true that these circumstances don't obtain if one is already assuming truthmaker necessitarianism! If necessitarianism is false, then the Xs might actually make it true that the circumstances are right even though they could exist and the circumstances be wrong. So Armstrong's proof of necessitarianism assumes what it aims to prove, and hence begs the question.¹¹

But nevertheless, even if we can't *prove* necessitarianism, I think it is a fairly attractive doctrine. The truthmaker thought is that explanation only bottoms out at existence facts: for God to give a complete plan of the world he need only make an inventory of what is to exist. But if necessitarianism is denied this doesn't seem to be the case; I can list everything there is and it still be an open question as to what is true. In determining what is to be the case, God can't just tell his angels what things to put in the world. He has to tell them also to make them a certain way. That goes against the whole spirit of truthmaker theory: explanation stops only at what there is—the ontological inventory being as it is should not leave the truth-value of any proposition undecided.

Necessitarianism is intuitive, and as a result I think we have a *pro tanto* reason to accept it. And so we should accept it, provided there is not an outweighing reason to reject it. Is there reason to reject necessitarianism? We can imagine two sorts of objection made to necessitarianism: (i) a proposed counterexample to necessitarianism, i.e. a case in which the Xs may plausibly be said to make $\langle p \rangle$ true but where the Xs could exist and $\langle p \rangle$ be false; or (ii) a rejection of the metaphysical assumptions underlying necessitarianism; for example a rejection of the claim that it makes any sense to talk of an object existing in counterfactual situations. Let's look at these in turn to see if there's reason to reject necessitarianism.

Counterexamples have been offered to necessitarianism by D. H. Mellor and John Heil, but I don't find them convincing. Mellor¹² asks us to consider a world, *W*, containing exactly two individuals,¹³ *a* and *b*, which are both *F*. What makes it true in *W* that every individual is *F*? Only this, says Mellor: the truthmakers for $\langle a \text{ is } F \rangle$ and for $\langle b \text{ is } F \rangle$. Those truthmakers do not *necessitate* that every individual is *F*, however; whatever they are, they could co-exist with some individual *c* which is not *F*. But the fact that there is no individual in *W* which is not *F* is a negative fact which, according to Mellor, does not itself require a truthmaker; negative facts, thinks Mellor, are not true because there is some truthmaker but are true because there's *not* a truthmaker for their negation. *a* and *b* are the only individuals in *W*; hence, thinks Mellor, $\langle \text{every individual is } F \rangle$ is true solely in virtue of that which makes *a* *F* and that which makes *b* *F*—there is no need to appeal to anything else.

¹¹ For more detailed discussion of this objection to Armstrong's argument see Cameron (2008b: 109–12).

¹² Mellor (2003).

¹³ Let's stipulate that properties and states of affairs are not individuals, just for ease of presentation.

Heil's proposed counterexample is as follows:¹⁴

Take the assertion

(P) If you drank this cyanide-laced tea, you would die.

Suppose (P) is true in virtue of some object or fact, a . . . Could we imagine a world that included a, but in which (P) was false?

Think of a world that included the cyanide-laced cup of tea but included, in addition, your having in hand an antidote. In that case, (P) could be false despite the presence of a . . . More generally, an assertion, A, might fail to hold, not because [the truthmaker for A] is absent, but because [the truthmaker for A] is accompanied by a defeater.

If we accept truthmaker maximalism—the doctrine that *every* truth has a truthmaker—these counterexamples are easy to respond to. Consider Mellor's. If every truth has a truthmaker then it seems that what really makes <every individual is F> true is not simply the truthmakers for <a is F> and <b is F> but these together with the truthmaker for <a and b are the only individuals>. But then the existence of *these* truthmakers necessitate the truth of <every individual is F>, and so it has a necessitating truthmaker after all.

Actually, that went a *bit* too quick. <Every individual is F> has a necessitating truthmaker *if* the truthmakers for the other three truths are necessitating truthmakers. But remember the dialectic: necessitarianism is intuitive; therefore we should accept it in the absence of reason to doubt it. So it's appropriate to assume that the truthmakers in question *other than that of the target proposition* are necessitating truthmakers. And if they are, then so is the truthmaker for <every individual is F> a necessitating one.

Similarly with Heil's example, if we are truthmaker maximalists then we will think that the truthmaker for (P) is not simply the dispositional properties of the cyanide, for example, but rather these *together* with whatever makes it true that there are none of the defeaters in question: i.e. it's not just the truthmaker for <cyanide is poisonous> that makes (P) true, but this together with the truthmakers for <you have no antidote to hand>, <there is not an exceptionally skilled doctor nearby>, <you do not have superhuman physiology>, and so on. As with Mellor's example, Heil's seems to trade on denying maximalism: the thought appears to be that the propositions proclaiming the *absence* of defeaters are not themselves made true by what there is, but are true because of what there's *not*—namely, defeaters. If we are maximalists there is no problem: (P) has a necessitating truthmaker—it is the necessitating truthmaker for <this cyanide is poisonous> taken together with the necessitating truthmakers for all the propositions proclaiming the absence of defeaters.

So if maximalism is true, the counterexamples to necessitarianism can be easily resisted. But if maximalism is false, I think they should also be resisted, because the falsity

¹⁴ Heil (2000: 232–3).

of maximalism should lead us to simply reject the claim that the target propositions have truthmakers at all, a fortiori that they have non-necessitating truthmakers.

In Mellor's case, why would one deny that that $\langle a$ and b are the only individuals that there are \rangle has a truthmaker while claiming that \langle everything is $F \rangle$ *does* have a truthmaker? Either totality facts need to be grounded in ontology or they don't. If they do, we resist the counterexample as above. If they don't, then *neither* of these propositions has a truthmaker; a fortiori, \langle everything is $F \rangle$ doesn't have a truthmaker whose existence fails to necessitate its truth, and so it doesn't provide a counterexample to necessitarianism.

Similarly with Heil's case, if maximalism is false then I think we should say that (P) itself doesn't have a truthmaker, a fortiori doesn't have a non-necessitating truthmaker. If you deny that claims like \langle there is no antidote \rangle have a truthmaker then such claims must be added to the class of brute truths; but now it simply becomes overwhelmingly plausible to say that (P) is true in virtue of not just the pure existence claims which proclaim the existence of whatever Heil takes to be the truthmaker for (P), but these claims together with the claims concerning the absence of defeaters that we were forced to add to the class of brute truths. In which case, by the definition of truthmaking I gave earlier, it will no longer be true to say that (P) has a truthmaker: its truth will not hold simply in virtue of some pure existence claim. And so, of course, we no longer have a counterexample to necessitarianism: (P) doesn't have a non-necessitating truthmaker—it doesn't have *any* truthmaker.

And I think this is going to generalize. If we assume maximalism we're never going to get a convincing counterexample to necessitarianism. We're only going to get a proposition that, intuitively, lacks a necessitating truthmaker if we assume that there's some other proposition that lacks *any* truthmaker. But such an assumption will simply undermine any reason we have for thinking that the target proposition has a truthmaker, thus undercutting the counterexample. And so we've seen no reason so far to suggest that a proposition can have a truthmaker whose existence nevertheless fails to necessitate the truth of that proposition.

What of the second type of reason one might mount against necessitarianism? Necessitarianism assumes that questions such as "Could the X s exist and $\langle p \rangle$ be false?" make sense and have answers. One might reject necessitarianism, then, either because one thinks that such questions are meaningless (perhaps because one is a Quinean abstentionalist about de re modality); or because one thinks that such questions do not admit of answers which are true or false (perhaps because one is a kind of expressivist about modality); or because one thinks that the answer to such questions will vary from context to context (perhaps because one is a counterpart theorist).

The latter two positions, however, don't offer a convincing reason to reject necessitarianism. Consider the case of the counterpart theorist. I take it that the argument against necessitarianism, for the counterpart theorist, would go something like this:

- 1) Whether or not $\langle p \rangle$ is made true by a is not a context-dependent matter.
- 2) Whether or not a could exist and $\langle p \rangle$ be false is a context-dependent matter.

- 3) Therefore, whether or not $\langle p \rangle$ is made true by a cannot depend on whether or not a could exist and $\langle p \rangle$ be false.
- 4) Therefore, it cannot be a condition on a 's being a truthmaker for $\langle p \rangle$ that the existence of a necessitate the truth of $\langle p \rangle$, and hence necessitarianism is false.

Now obviously one could reject premise (1). That was Lewis's preference.¹⁵ The truthmaker demand is to find an object for every truth whose existence necessitates that truth. And according to Lewis, whether it's true to say that A 's existence necessitates the truth of p depends on whether A has a counterpart in a world in which p is false, which depends on what things in what worlds get to count as A 's counterparts, which depends on which of A 's properties are contextually salient, which depends on how we're thinking about A . So, at least in many cases, we can introduce a context in which it is true to say that A 's existence necessitates p simply by referring to A in a way that makes the relevant properties salient. Consider for example the truth $\langle \text{Socrates is bearded} \rangle$. Typically, truthmaker theorists think Socrates himself couldn't make that true, since he could have lacked a beard. But for the counterpart theorist, that's just to say that there are possible beardless beings who are similar enough to Socrates in salient respects. So all we need to do is to make Socrates' beardedness salient; by referring to Socrates as "Socrates-qua-bearded" we invoke a counterpart relation according to which no beardless being is similar enough to Socrates to be his counterpart. Thus "Socrates-qua-bearded couldn't exist and $\langle \text{Socrates is bearded} \rangle$ be false" is true. But "Socrates-qua-bearded" just refers to Socrates. So, says Lewis, Socrates is (in this context) an appropriate truthmaker for the fact that he is bearded.

There's much to say about Lewis's truthmaker theory.¹⁶ For starters, given our above rejection of necessitation as *sufficient* for truthmaking, one might simply deny that Socrates-qua-bearded makes it true that Socrates is bearded. Necessitation is not enough: $\langle \text{Socrates is bearded} \rangle$ must be true *in virtue of* the existence of the truthmaker, and it is not true in virtue of Socrates, no matter what properties of Socrates are contextually salient. Once we find the real truthmaker—the state of affairs of Socrates being bearded, perhaps—the counterpart theorist might hold that there's no context in which one can truly say that this thing might have existed and Socrates not been bearded. That would be to be less liberal about counterpart theory than Lewis, who apparently thinks that for any possibly true proposition p , and for any thing A , there's a possible context in which there's a counterpart of A in a world where p is true—but there is nothing about counterpart theory per se that forces one to be this liberal.

But even the completely liberal counterpart theorist can resist the above argument if she wishes, and hence hold on to necessitarianism. The inconstancy of modality *de re* together with necessitarianism does not mandate the inconstancy of facts concerning what grounds what. The inconstancy of modality *de re* arises, for the counterpart theorist,

¹⁵ Lewis (2003).

¹⁶ See Cameron (2008c: 29–31) and MacBride (2005) for discussion.

because what propositions are expressed by de re modal claims, and what properties are the semantic values of de re modal predicates, varies from context to context; it does *not* arise because the propositions expressed vary their truth value from context to context, or because an object changes its modal properties from context to context. If “*a* could have been F” expresses a truth in a context C_1 and a falsehood in context C_2 then that is not because there is one proposition that is expressed in both contexts and that is true in C_1 but false in C_2 , nor is it because *a* has different modal properties relative to each context; rather, it is because there is a proposition that is true *simpliciter* that is expressed by that sentence when uttered in context C_1 and a proposition that is false *simpliciter* that is expressed by that sentence when uttered in context C_2 . *a* could have been F iff it has a counterpart that is F; so when I say “*a* could have been F” I am saying that *a* has a counterpart that is F. But if this sentence expresses a truth in C_1 but a falsehood in C_2 then that is because a different counterpart relation is invoked from one context to another as a result of different standards of similarity being contextually salient; so in uttering that sentence in C_1 I express the true (*simpliciter*) proposition that there is a being that is F that bears the counterpart₁ relation to *a*, and in uttering the same sentence in C_2 I express the false (*simpliciter*) proposition that there is a being that is F that bears the counterpart₂ relation to *a*.

Now, if one thought that the de re modal facts changed their truth-value from context to context, and one didn’t want to hold that the facts concerning what grounded what so changed their truth-value, then maybe one would indeed have a reason to deny necessitarianism (or at least, to claim that “necessitarianism is true” does not express a truth in every context). But that’s not what the counterpart theorist thinks; and indeed, it’s not what one should think. And the actual reason for the inconstancy of modality de re (assuming it *is* inconstant) gives us no reason to deny necessitarianism or to accept the inconstancy of grounding. For there may be a unique counterpart relation that is invoked in contexts where what is salient is that we are engaged in discussions of truthmaking. Call this the truthmaking counterpart relation. In stating truthmaker necessitarianism we create a context in which this is the counterpart relation that is invoked, so truthmaker necessitarianism says that if *a* is a truthmaker for $\langle p \rangle$ then every world in which *a* has a counterpart under the truthmaking counterpart relation is a world in which $\langle p \rangle$ is true. That, if true, will be true *simpliciter*, and not a context-dependent matter. The fact that we can invoke a context whereby we can truly say that *a* could exist and $\langle p \rangle$ be false does not imply that there is a context whereby *a* is not a truthmaker for $\langle p \rangle$, for the counterpart relation that is invoked in the context under consideration is not the truthmaking counterpart relation, and so whether or not there is a world at which there is a counterpart of *a* according to this new counterpart relation and at which $\langle p \rangle$ is false is simply irrelevant to whether or not $\langle p \rangle$ is made true by *a*.

The moral is that the status we accord to de re modal judgments needn’t be the same as the status we accord to judgments concerning what makes what true, even if we accept truthmaker necessitarianism. Inconstancy de re does not imply the inconstancy of grounding. For similar reasons, one need not reject truthmaker necessitarianism on the

grounds that *de re* modal claims are not truth-apt but claims about truthmaking are, even if one believes both those claims. If I am an expressivist about modality and hold, for example, that “*a* could be red” doesn’t aim to state a truth but is rather an expression of my finding intelligible a scenario in which *a* is red, it is nevertheless perfectly consistent for me to hold that “*a* makes $\langle p \rangle$ true” *does* express a truth *and* to hold that *a* makes $\langle p \rangle$ true only if *a* couldn’t exist and $\langle p \rangle$ were false: the reason I hold to necessitarianism may simply be that what I find intelligible is constrained by my beliefs concerning the facts about grounding, so that my believing that *a* makes $\langle p \rangle$ true renders it unintelligible to me that *a* could exist and $\langle p \rangle$ be false, thus explaining my utterance of “*a* couldn’t exist and $\langle p \rangle$ fail to be true.”

The constraints that truthmaker necessitarianism place on one’s theory of modality are not as great as one might initially have thought, then. There are some constraints, to be sure: the necessitarian has to admit that *de re* modal talk makes sense, and she has to allow that there are some true, or at least assertable, essentialist claims, lest she weaken her truthmaker theory to the point of uselessness. (An extreme anti-essentialist who thought that for all objects *x*, “*x* couldn’t exist and $\langle p \rangle$ be false” was only true/assertable if $\langle p \rangle$ couldn’t be false could obviously only hold to necessitarianism if she denied that any contingent truth had a truthmaker.) But these are fairly minimal commitments—they are compatible with all manner of views about the status of *de re* modal claims—and so I don’t see much of a threat to necessitarianism here.

13.4 TRUTHMAKER MAXIMALISM

I said that truthmaker theory at its most ambitious is the conjunction of two claims:

Priority of Existence: The brute truths are a subset of the pure existence claims

Universal Grounding: Every true proposition is either brute or is true in virtue of some brute proposition

Acceptance of Priority of Existence and Universal Grounding commits one to truthmaker maximalism—the doctrine that *every* truth has some truthmaker(s). This is a controversial thesis; Peter Simons calls it “the most tendentious of Armstrong’s general truthmaking principles.”¹⁷

What is the source of suspicion against maximalism? The biggest source of resistance to maximalism stems from a suspicion of truthmakers for negative truths: facts about what is *not* the case. Negative truths, many think, are true not because of what there *is* but because of what there *isn’t*: \langle It’s not the case that electron *e* is positively charged \rangle is true not because *it* has a truthmaker but because there is *no* truthmaker for $\langle e$ is positively

¹⁷ Simons (2005: 254).

charged>. Particularly salient are negative existentials: truths concerning what there isn't, such as <there are no unicorns>. The suspicion against positing truthmakers for such claims is well summed up by Joseph Melia:¹⁸

It is simply wrong-headed to think that there should be a truthmaker which corresponds to negative existential sentences. Intuitively, what makes a sentence such as "there are no Fs" [true] is a *lack* of Fs . . . it is just confused to think that we must account for a lack by postulating the *existence* of something else.

I think Melia is correct that, *intuitively*, negative existentials are not true because of what there is but because of what there isn't. If so, that's a pro tanto reason to reject truthmaker maximalism, and I accept that. The question then is whether there's good reason to accept maximalism, and whether that outweighs the intuitive reason to reject it. We'll come back to this.

Another argument against positing truthmakers for negative truths comes from George Molnar, who put the problem in terms of four propositions which are in apparent tension:¹⁹

- (1) The world is everything that exists
- (2) Everything that exists is positive
- (3) Some negative claims about the world are true
- (4) Every true claim about the world is made true by something that exists

(3) and (4) tell you that some negative claim about the world is made true by something that exists; (2) tells you that this thing is positive; hence, a negative claim about the world is made true by a positive thing. This is meant to sound unintuitive: how can a positive thing make true a negative claim? So since (1)–(3) are meant to sound highly intuitive, we should reject (4). We could modify (4) to yield the conclusion that the negative claim is made true by something that *doesn't* exist; but then (1) would force us to deny that the negative claim is made true by the world. So, we conclude, negative claims are not made true at all.

I am with Josh Parsons²⁰ in not finding this convincing due to dissatisfaction with (2). What is it for a *thing* to be positive or negative? Like Parsons, I have no idea. Being positive or negative seems to apply, in the first case, to representational entities such as propositions: <there is a donkey> is positive and <there is not a talking donkey> is negative because the former represents something to be the case and the latter represents something *not* to be the case. But most things are not representations, so it seems that we can call them "positive" or "negative" in a derivative sense at best. So what are the rules? Is an entity positive if it corresponds to a positive proposition and negative otherwise?

¹⁸ Melia (2005: 69).

¹⁹ Molnar (2000: 72).

²⁰ Parsons (2006).

If so, then we should not be convinced by Molnar's (2). (2) would amount to the claim that nothing corresponds to a negative claim; but the correspondent of a truth, presumably, is just its truthmaker, so (2) would simply amount to the claim that there are no truthmakers for negative claims. It can hardly be relied upon, then, in an argument meant to show that there's a problem in providing truthmakers for negative truths. But if that's not what it means for an entity to be negative or positive then I have no idea what it means. So Molnar doesn't seem to have provided any particular reason for thinking that negative truths resist truthmaking.

I think there is no reason to simply rule out truthmakers for negative truths from the get-go. But admitting that negative truths have truthmakers *does* require us to admit some rather strange things into the world. What makes it true, for example, that there are no unicorns? Truthmaker theorists disagree, but options include: (1) the absence of unicorns;²¹ (2) the totality fact that all the first-order states of affairs are *all* the first-order states of affairs;²² (3) the world itself, which by its very essence contains no unicorns.²³ It's fair to say that each option requires us to believe that ontology is significantly richer than we might otherwise have thought, and so one who denies that negative truths have truthmakers can claim to have a simpler ontology. In that case we should ask, whether the admission of these truthmakers is *worth* the extra ontological expense? But there's nothing special about negative truths here: this is the situation we end up in for most domains of truths. Setting aside the most mundane of truths, admitting truthmakers for a domain of truths requires us to complicate our ontology in ways we otherwise wouldn't have to. Consider simple intrinsic predications like <Frodo is brave>. Such claims are not in general necessary if true; Frodo is *in fact* brave, but he surely might have been cowardly. In that case, given truthmaker necessitarianism, Frodo can't himself be the truthmaker for <Frodo is brave>. So truthmaker theorists are driven to complicate their ontology by admitting something the denier of truthmaker theory doesn't have to believe in, such as the state of affairs of Frodo instantiating the property of being brave, or the particular braveness that can only be had by Frodo. As always, the question comes down to whether these ontological posits are worth the benefit they bring. Truthmakers for negative truths are no different. So *is* the admission of these truthmakers worth the cost? To answer that question we must first know what the benefit of having them is: why be a truthmaker theorist in the first place?

13.5 REASONS TO BE A TRUTHMAKER THEORIST: 1, 2, 3

I'll look at three reasons to be a truthmaker theorist. Each, if sound, would give us a reason to accept that negative truths have truthmakers as well as positive ones. I'll argue

²¹ Martin (1996).

²² Armstrong (1997; 2004).

²³ Cameron (2008d).

that the first two reasons shouldn't move us, but that the third gives us at least *pro tanto* reason to be truthmaker theorists.

13.5.1 Truthmakers are required by the theory of truth

If we accept a theory of truth whereby *what it is* for a truth to be true is for it to be made true then this should lead us to accept that *all* truths, a fortiori negative truths, have truthmakers.

Here's my earlier temporal part:²⁴

Truthmaker theory is a theory about *what it is* for a proposition to be true; it's just not the kind of theory that can apply only in a restricted domain. What possible reason could one have for thinking of some propositions that they need to be grounded in what there is that doesn't apply to all propositions? Why should it be okay for negative truths to go ungrounded and not okay for positive truths to go ungrounded?

This echoed Armstrong,²⁵ who argues that one who believes that some but not all truths have truthmakers thereby commits to an unattractive dualism about truth, thinking there are two ways for propositions to be true: some are true by corresponding to portions of reality, some are true in some other manner.

I now think that this is a mistake. Truthmaker theory should *not* be construed as a thesis about *what it is* for a proposition to be true. Were it so construed, there is simply no good reason to accept it.²⁶ Truthmaker theory should be solely a thesis concerning what truths are, *as a matter of fact*, brute. Nothing about the *nature* of truth justifies our acceptance of truthmaker theory. Our explanation for *why* a proposition is true and our theory of *what it is* for that proposition to be true are two separate things, and they need not be related. The truthmaker theorist says that every truth is true *because* of what there is, and this is simply silent on the question of *what it is* for a truth to be true in the first place. It's compatible with every truth being *made* true that there's nothing more to a proposition's *being* true than that what it says is the case *is* the case. Truthmaker theory is compatible with a deflationist theory of truth.²⁷

²⁴ Cameron (2008d: 412).

²⁵ Armstrong (2006: 245).

²⁶ Perhaps certain versions of the correspondence theory of truth will entail the truthmaker principle; but I would like a justification for truthmaker theory that appealed to more than those who accept the correspondence theory of truth. See David (ch. 9 in this volume) for a discussion of the correspondence theory.

²⁷ For a discussion of deflationism, see Azzouni (ch. 17 in this volume).

13.5.2 Truthmaker theory is required by realism

Some think that truthmaker theory is a commitment of realism. John Bigelow says,²⁸

I have sometimes tried to stop believing in the Truthmaker axiom. Yet I have never really succeeded. Without some such axiom, I find I have no adequate anchor to hold me from drifting onto the shoals of some sort of pragmatism or idealism. And that is altogether uncongenial to me; I am a congenital realist about almost everything.

And Armstrong says,²⁹

My hope is that philosophers of realist inclinations will be immediately attracted to the idea that a truth, any truth, should depend for its truth [on] something “outside” it, in virtue of which it is true.

And Heil calls the truthmaker principle “a central tenet of realism.”³⁰ If truthmaker theory is required by realism then we must admit truthmakers for at least all those truths the domain of which we want to be realists about. So you might hold that <Allegri’s *Miserere* is beautiful> is true but lacks a truthmaker if you are a certain kind of anti-realist about beauty; but presumably at least some negative truths are as really true as anything is, so this reason to accept truthmaker theory would move us to admitting truthmakers for some negative truths if it moved us at all. But again, I don’t think we should be moved by this reason. I simply can’t see why one should think that realism requires truthmakers. Realism is to do with objectivity, mind-independence, evidence-transcendence, and so on.³¹ What does any of that have to do with there being some things for each truth, making that truth true? Certainly, being a truthmaker theorist doesn’t seem to be *sufficient* to be a realist: you could think that the truthmakers were all ideas in the mind of God. To be a *realist* truthmaker theorist you’d have to insist that the truthmakers were mind-independent entities, or objective, or the subjects of evidence-transcendent truths, or something like that. But then it seems like it’s the mind-independence/objectivity, etc. that’s doing the work, not the truthmaking: if our only concern is being a realist, why not just insist that the *truths* are mind-independent/objective, etc.—what work does going via truthmakers do? In any case, the onus is on those who posit a connection between truthmaker theory and realism to argue for it; but in my opinion, they have yet to do so.

²⁸ Bigelow (1988: 123).

²⁹ Armstrong (2004, 7).

³⁰ Heil (2003: 61).

³¹ For discussions of realism see Shieh (ch. 16 in this volume).

13.5.3 Truthmaker theory yields a small class of brute truths

We've looked at two attempts to motivate truthmaker theory and found them wanting. Let me now present what motivates me at least to accept truthmaker theory. If the notion of metaphysical explanation makes sense then *some* class of truths must be the brute truths. If it is to be a useful notion then it must be a proper sub-class of the class of all truths. The question, then, is which, out of all the truths, are brute? Might none be brute? Might the class of brute truths be the empty set? There are good arguments that suggest not.³² In that case, the task becomes to locate the non-empty proper sub-class of all the truths that we should consider to provide the grounding for all the rest of the truths.

I propose the following as a good methodological principle: keep the class of brute truths minimal in the sense both that it contains as few truths as possible and that it contains as few *kinds* of truth as possible.³³

The idea is that we should minimize the amount of bruteness that our theory implies. If there are two theories, T₁ and T₂, which disagree over what the brute truths are, and if T₁ says that the brute truths are a proper subset of the truths T₂ says are brute, then this is a pro tanto benefit in favor of T₁ over T₂. Similarly, if T₁ says the brute truths are exhausted by the truths of kind K₁ and K₂ but T₂ says that some truths of kind K₃ are brute *as well*, then this is a pro tanto benefit in favor of T₁ over T₂. These methodological principles are motivated by parsimony considerations. It's better to ground truths than to leave them brute: brute truth is a cost against a theory. There must be *some* brute truths, but we should have as few as possible, of as few kinds as possible.

If this methodological constraint is sound then truthmaker theory is off to a good start, for it identifies as the brute truths a relatively sparse set of propositions that are of a kind. While some people want to take as brute what there is and *how* those things are, or what there is and how those things are and how they *were* and how they *will* be, or all this together as well with how they *could* be, the truthmaker theorist thinks that the only brute truths are those concerning what there *is*; thus they gain a benefit over their more expansive rivals. Now, I'm certainly not saying that this rules conclusively in favor of truthmaker theory: all I'm suggesting is that there is a pro tanto case for truthmaker theory. If everything we want grounded can be grounded in the sparse base the truthmaker theorist allows herself then, other things being equal, we should accept truthmaker theory. Of course, other things are not equal: truthmaker theory requires complicating our ontology. The question is whether such ontological complications are worth the benefit in minimizing bruteness. To the victor, the spoils!³⁴

³² See Hudson (1997) for discussion.

³³ It's possible, of course, that the two senses of minimality could pull against one another, with a reduction in the number of kinds of truth taken as brute achievable only via an increase in the number of truths taken as brute, for example; but the principle is none the worse for that.

³⁴ For a more detailed presentation of this defence, see Cameron (2008b). Cf. Nolan (2008), who can be seen as arguing that truthmaker theory is beneficial because it allows us to not take facts about instantiation as brute.

The proper justification for maximalism, then, is simply that it is beneficial if true. Any departure from maximalism requires some addition in the truths that are taken as brute. The maximalist's brute truths will all be of the form "... exists," whereas the non-maximalist will *also* have to take as brute truths which do not simply characterize the ontological inventory. So for example a presentist who denies that historical truths are made true by present entities must simply take as brute facts about what *was* the case;³⁵ one who doesn't admit something like a totality fact will have to take as brute facts about what *doesn't* exist. That, I suggest, is a cost to the theory, thus making truthmaker theory advantageous in this respect.

Sometimes non-maximalists try to sound like maximalists. Jonathan Tallant, a presentist, rejects the demand for present truthmakers for historical facts such as <Socrates existed>, claiming instead that we should demand merely that there *did* exist an entity (in this case, Socrates) that makes the historical fact true.³⁶ Compare the quote earlier from Melia, to the effect that, intuitively, <There are no Fs> is made true by the *lack* of Fs. This makes it sound like these propositions *are* being made true; it's just that they're being made true by what there *was* or what there's *not*, as opposed to being made true by what there *is*.

But consider Tallant's claim: there *did* exist an entity that makes <Socrates existed> true. It's crucial that we're using "makes" here, rather than "made": we're interested in what accounts for the present truth of <Socrates existed> rather than what accounted for the truth of <Socrates exists> in the past. For a presentist to claim that this is now made true by an entity that *did* exist is to claim that non-existent objects nonetheless presently act as truthmakers. I don't understand how something can be a truthmaker for a truth without existing!

In my view, saying that historical facts or negative existentials, etc. are made true by what there was or what there's not, etc. is simply disingenuous when one doesn't believe in absences, or things that don't exist, or in past ontology, etc. Better for the non-maximalist to just come clean and say what they really commit themselves to: that these claims are simply not grounded by anything! To say that <there are no Fs> is made true by the absence of Fs, when you don't believe that there is some thing that is the absence of Fs, is just to say that <there are no Fs> is true because there are no Fs. Well, "<p> is true because p" might be an explanation in *some* sense; but if that's the end of the story then p must be a brute truth: and so if p is a historical fact, or a modal fact, or a negative existential, then this involves enriching the class of brute truths to include facts of a kind other than what there is. While the truthmaker theorist's brute truths are all statable using only the existential quantifier, names, and an identity predicate, her opponent has to allow herself primitive negation, or primitive tense, or primitive modal operators, etc., in order to state the brute truths of the world.

³⁵ Whether the presentist can provide presently existing truthmakers for historical facts—and whether, if not, it counts against presentism or against truthmaker theory—is a vexed issue. See Sider (2001: ch. 2) for the truthmaker objection to presentism and Merricks (2007: ch. 6) for the presentist objection to truthmaker theory; see Cameron (2011) for an attempt at reconciliation.

³⁶ Tallant (2009).

At this point, then, I claim that we need to balance up the pros and cons: deniers of maximalism have a smaller ontology, but the maximalist has a sparser class of brute truths. But Tallant thinks that Ockham's razor *mandates* us to accept the non-maximalist's account over the maximalist's. He says,³⁷

Ockhamistic considerations require us not to multiply entities beyond necessity . . . I've suggested, we could make our statement of when propositions are true, more complex, rather than positing additional ontology. Thus, there is no *necessity* to introduce further entities.

The thought here seems to be that because it's *possible* to have a richer class of brute truths and thus avoid additional ontology, admitting the truthmakers as ontological grounds is ruled out by Ockham's razor. I think this is wrong. Ockham's razor had better allow us to say that ontological parsimony can be outweighed, otherwise we shouldn't believe in *anything at all!*³⁸ That is, that we're only *mandated* to accept the ontologically sparse theory if it can be obtained without incurring some *additional* cost. But of course, once this concession is made, Ockham's razor won't rule out truthmaker maximalism because, as argued above, its rivals *do* incur extra costs: the admission of more brute truths. So we're back to weighing up the pros of minimizing brutality against the cons of ontological extravagance.

Every theory should tell us three things: what exists, what is true, and what the relationship is between what is true and what exists. It's extremely easy to be minimal about what exists, if we don't care about saying implausible things in the other two cases: if we're either willing to deny lots of truths (e.g. denying that there are any true claims about the past), or to say that there's just no link between truth and ontology,³⁹ then we can have as sparse an ontology as we like. Similarly, it's extremely easy to have a simple theory of how truth is grounded in ontology (namely, the theory that every truth has a necessitating truthmaker) if we either don't care about having a minimal or plausible ontology (we can for instance just postulate the fact that p, for any true proposition p) or are again willing to deny the truth of many claims (e.g. if we held that the only truths were truths about what there is). Ockham's razor should partly be understood as telling us not to answer the three demands such that something in your posited ontology isn't, by the lights of your principle concerning how truth depends on the world, doing any work in accounting for the claims your theory says is true. Ockham's razor, thus understood, is an internal monitor on theories: it tells you not to believe in things that don't do any work by your own lights in accounting for what you think is true. What Ockham's razor *won't* do is tell you to believe that historical facts are brute and deny past entities

³⁷ Tallant (2009: 427). Punctuation as in the original.

³⁸ For a defense of the claim that it's coherent to believe ontological nihilism—the view that there is nothing at all—see Turner (2011). The reason not to be an ontological nihilist is that whilst it has a minimal ontology, it is costly in other respects: precisely the kind of reason Tallant's version of Ockham's razor is blind to.

³⁹ As Turner's Propositional Ontological Nihilist does (Turner 2011).

rather than accepting past entities, or present past-directed truthmakers, and believing truthmaker theory. To decide between these theories we need to look at the answers they give to all three of the above demands, and evaluate their joint attractiveness.

In general, giving an attractive answer to one of the three demands is easy. What's hard is giving an attractive answer to all three. And that's where the fun lies: the challenge facing us is to give an answer to all three that is better than any alternative *combination* of answers. We shouldn't be worried if rival theories give a better answer to *one* demand: after all, we should expect these demands to pull against one another—that giving a nice answer to one is going to be paid for by giving up some of the nicer answers to another. We're only *mandated* to reject our theory and accept a rival if it answers one of the demands better than we do and answers no demand worse; otherwise, we must simply engage in the holistic weighing-up of costs against benefits—and there will be no obvious answers.

Truthmaker theory offers an attractive account of how truth depends on ontology, and it is non-revisionary with respect to what we pre-philosophically believe. It manages this, admittedly, at the expense of unusual ontology. So the challenge for truthmaker theorists is to make their ontology as acceptable as possible and thus minimize their costs, and the challenge for their opponents is to show that the benefit is not worth even these costs.

REFERENCES

- Armstrong, D. (1997). *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Armstrong, D. (2004). *Truth and Truthmakers*. Cambridge: Cambridge University Press.
- Armstrong, D. (2006). "Reply to Heil," *Australasian Journal of Philosophy* 84: 245–7.
- Bigelow, J. (1988), *The Reality of Numbers: A Physicalist's Philosophy of Mathematics*. Oxford: Clarendon Press.
- Cameron, R. (2008a). "Truthmakers and Ontological Commitment: Or, How to deal with Complex Objects and Mathematical Ontology Without Getting into Trouble," *Philosophical Studies* 140: 1–18.
- Cameron, R. (2008b). "Truthmakers, Realism and Ontology," in R. LePoidevin (ed.), *Being: Contemporary Developments in Metaphysics*. Royal Institute of Philosophy Supplement 83, 107–28.
- Cameron, R. (2008c). "Truthmakers and Necessary Connections," *Synthese* 161: 27–45.
- Cameron, R. (2008d). "How to be a Truthmaker Maximalist," *Noûs* 42: 410–21.
- Cameron, R. (2011). "Truthmaking for Presentists," in K. Bennett and D. W. Zimmerman (eds.), *Oxford Studies in Metaphysics*, vol. 6. Oxford: Oxford University Press, 55–100.
- Heil, J. (2000). "Truthmaking and Entailment," *Logique et Analyse* 43, 231–42.
- Heil, J. (2003). *From an Ontological Point of View*. Oxford: Oxford University Press.
- Hudson, H. (2007). "Brute Facts," *Australasian Journal of Philosophy* 75: 77–82.
- Lewis, D. (2003). "Things qua Truthmakers," in H. Lillehammer and G. Rodriguez-Pereyra (eds.), *Real Metaphysics*. London: Routledge, 25–38.
- MacBride, F. (2005). "Lewis's Animadversions on the Truthmaker Principle," in H. Beebe and H. Dodd (eds.), *Truthmakers: The Contemporary Debate*. Oxford: Oxford University Press, 117–40.

- Martin, C. B. (1996). "How It Is: Entities, Absences and Voids," *Australasian Journal of Philosophy* 74: 57–65.
- Melia, J. (2005). "Truthmaking without Truthmakers," in H. Beebe and J. Dodd (eds.), *Truthmakers: The Contemporary Debate*. Oxford: Oxford University Press, 67–84.
- Mellor, D. H. (2003). "Real Metaphysics: Replies," in H. Lillehammer and G. Rodriguez-Pereyra (eds.), *Real Metaphysics*. London: Routledge, 212–38.
- Merricks, T. (2007). *Truth and Ontology*. Oxford: Clarendon Press.
- Molnar, G. (2000). "Truthmakers for Negative Truths," *Australasian Journal of Philosophy* 78: 72–86.
- Nolan, D. (2008). "Truthmakers and Predication," in D. Zimmerman (ed.), *Oxford Studies in Metaphysics*, vol. 4, 171–92.
- Parsons, J. (2006). "Negative Truths From Positive Facts?" *Australasian Journal of Philosophy* 84: 591–602.
- Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press.
- Simons, P. (2005). "Negatives, Numbers and Necessity: Some Worries about Armstrong's Version of Truthmaking," *Australasian Journal of Philosophy* 83: 253–61.
- Tallant, J. (2009). "Ontological Cheats Might Just Prosper," *Analysis* 69: 422–30.
- Turner, J. (2011). "Ontological Nihilism," in K. Bennett and D. W. Zimmerman (eds.), *Oxford Studies in Metaphysics*, vol. 6. Oxford: Oxford University Press, 3–54.

CHAPTER 14

A LOGICAL THEORY OF TRUTH-MAKERS AND FALSITY-MAKERS

NEIL TENNANT

14.1 INTRODUCTION

DUMMETT's regulative principle C (for "Correspondence," stated in Dummett 1976: 89) is that

If a statement is true, there must be something in virtue of which it is true.

The thing in question is aptly called a truth-maker. To date, no one really knows what such a thing may be.

In the 1980s three seminal publications appeared, which put truth-makers and truth-making at center stage. These were the 1984 paper "Truth-Makers," by Kevin Mulligan, Peter Simons, and Barry Smith; the 1987 paper "Truthmaker" by John Fox; and the 1988 discussion of truth-makers by John Bigelow in Part III of his book *The Reality of Numbers*. The field is now well served by the two recent anthologies Beebe and Dodd (2005) and Lowe and Rami (2009), each with a thorough and competent editorial introduction that surveys the issues involved. This happy situation moves the present author to avoid re-hashing the material with a survey-style article, and rather to go back to original sources, reflect on their materials, and then attempt to strike out in a new direction that is informed by the author's own methodological interests. We commend to the reader a new, *logical* theory of truth-makers, a theory that deals with *logically complex* statements as well as logically primitive, or atomic, ones.

14.2 THE NOTION OF A TRUTH-MAKER

Here is how the aforementioned authors of the three seminal publications interpreted the notion of a truth-maker.

... entities *in virtue of which* sentences and/or propositions are true ... we shall ... [call] ... *truthmakers*. (Mulligan et al. 1984: 287–8)

By the truthmaker axiom I mean the axiom that for every truth there is a truthmaker; by a truthmaker for A, I mean something whose very existence entails A. (Fox 1987: 189)

... whenever something is true, there must be something in the world which makes it true. I will call this the Truthmaker axiom ... A truthmaker is simply an object whose existence entails a truth. (Bigelow 1988: 122 and 128)

These three treatments of truth-makers venture to characterize truth-makers only for atomic sentences or propositions; and the main monograph-length treatment, Armstrong (2004), at best succeeds in characterizing truth-makers only in the atomic case.¹ It would seem that for logically complex sentences (or propositions)²—most notoriously, negations, disjunctions, and universal quantifications—truth-makers are very hard to come by. Such, at least, is the lesson these recent pioneers appear to draw from the tradition:

The glory of logical atomism was that it showed that not every kind of sentence needs its own characteristic kind of truth-maker. Provided we can account for the truth and falsehood of atomic sentences, we can dispense with special truth-makers for, e.g., negative, conjunctive, disjunctive, and identity sentences. (Mulligan et al. 1984: 288)

... what could be the truth-makers ... for universal claims, such as that all ravens are black? The sum of the truthmakers for all the truths of the form “a is black” would not do, for they could exist yet one extra raven exist and be purple. What is required beyond this is that there be no ravens but these; and what would be the truthmaker for *this*?

One strategy for dealing with this is to abandon the generality of the axiom by restricting it to atomic claims.

Such explicit restriction raises the problem of explaining the relation of truth in general to truthmakers ... [A]rguably it is only atomic truthmaker that can cogently be attributed to the tradition. (Fox 1987: 204–5)

¹ Armstrong has written to the author, “... I myself would reject the idea that truthmakers should just be given for “atomic” truths. I’m a truthmaker maximalist—every truth has a truthmaker. I’m not sure, of course, for many truths, what their truthmakers are.”

² Henceforth we shall concentrate on just sentences, pending an interesting, theoretically-motivated twist that will emerge in due course.

14.3 MATTERS OF BACKGROUND

Some truth-making theorists—especially the Australian school—take themselves to be indulging an unrelenting realism in metaphysical outlook, in hypothesizing that every truth has a truth-maker. But to Dummettian anti-realists there is nothing untoward about this truth-maker axiom, *given the anti-realist's conception of what truth consists in*. Truth, for the anti-realist, consists in the existence of a truth-maker, albeit a truth-maker of a very special kind—one that is in principle surveyable, or epistemically accessible. For the anti-realist, all truths are knowable. And that is why truth cannot be relied upon to be bivalent.

It is an old point (due to Putnam), but one worth re-emphasizing here, that Tarski's theory of truth, including its inductive definition of satisfaction and its derivations of all instances of the *T*-schema, are available to the anti-realist even after the anti-realist has restricted herself to intuitionistic logic. (The present author would strengthen this observation: Tarskian truth theory can be carried out within the yet more restricted system of *core logic*, which results from classical logic by pursuing not only constructivizing reforms, but relevantizing ones also.³)

The only "classical" or "realist" feature of Tarski's theory of truth obtrudes at the point where the theorist tries to prove (in the metalanguage) the Principle of Bivalence: that for every sentence ϕ of the object language, either ϕ is true or $\neg\phi$ is true. The proof of *that* metalinguistic result requires the use, in the metalogic, of a strictly classical rule of negation, such as the law of excluded middle, or the rule of dilemma. But by contrast the derivations of all instances of the *T*-schema proceed in innocence of such classical rules. The derivations are available in core logic.

So the anti-realist—even one who uses only core logic—can hear the truth-making theorist's asseveration that every truth has a truth-maker, and unequivocally agree.

It follows that if the truth-making theorist wishes to express his metaphysical *realism* by postulating the existence of truth-makers for each and every truth, he needs a more exigent account of what truth-makers are. He needs to characterize them in a way that enables one subsequently to argue, cogently, that for every sentence ϕ , either there is a truth-maker for ϕ or there is a truth-maker for $\neg\phi$.

The present author is about to share with his realist opponents one possible way of doing this. But we shall be careful to point out later (in section 14.11.4) where exactly, in the account about to be given, the anti-realist should demur.

³ Core logic is what the present author called *intuitionistic relevant logic* in publications such as Tennant (1987) and Tennant (1997). The new label is preferable in light of an argument in Tennant (unpublished typescript, 2012) for an important revision-theoretic thesis:

Core logic is the minimal inviolable core of logic without any part of which one would not be able to establish the rationality of belief-revision.

14.4 THE AIMS OF THE PRESENT STUDY

In this chapter we seek to make good the above-mentioned omission on the part of both vintage and recent truth-maker theorizing, even though it is in realist vein. We offer a new explication of the *kind of things* truth-makers are. The aim is to allow the claim, in full generality, that every truth has a truth-maker.⁴ Ours is a *structural* theory of truth-makers. They are not (to take one extreme) simply individuals, or *Sachverhalte* composed of individuals. Nor, at the other extreme, are they the Great Fact.

Rather, our truth-makers are *proof-like* objects—possibly infinitary, depending on the size of the domain—that clearly *articulate* the grounds for truth of the claim in question.⁵

An immediate note of caution should be entered here. Throughout this discussion, unless otherwise indicated, the phrase “grounds for truth” and its cognates are to be construed *metaphysically* or *ontologically*, rather than epistemically. We are thinking of a species of ontological grounding here, or of constitution—of the actual making-true. It would be a serious error to read the phrase “grounds for truth” as evidential or epistemic in connotation—as having to do with, say, grounds for belief or justification. There might be an interesting relationship, to be explored, between the two kinds of ground; but nothing we say subsequently will be properly understood if we are mistakenly construed, from this point on, as talking about evidential grounds when in fact we are talking about constitutive ones.⁶

Our truth-makers will be able to contain actual individuals. Individuals will be able to be embedded within so-called *saturated formulae* (see section 14.5.1) that occur within truth-makers, such occurrences being like those of ordinary (formal) sentences within deductive proofs (which we shall cast as natural deductions). Moreover, just as a theorem of a mathematical theory can enjoy more than one proof from the axioms of the theory, so too will a given sentence be able to enjoy more than one truth-maker (relative to a given model, or interpretation). One’s immediate intuition is that this is how it ought to be—for in general there are more ways than one for a claim to be true (even within the context of an interpretation, or model, held fixed).

We stress again that our truth-makers will actually *embody the individuals involved*, embedded within what we are calling *saturated formulae*. So, abstract though they may still be, the truth-makers will “limn the lines of factual composition,” as it were, beginning with the truth (or falsity) of basic predications embedding those individuals. It is difficult to imagine any other way of formulating a theory of truth-makers that can hope

⁴ This claim is intended to survive even in the context of *semantically closed* languages, which are plagued by the logical and semantic paradoxes. See Tennant (2012) for details.

⁵ Sundholm (1994) points out that *for the intuitionist* truth-makers are finitary proofs. He does not, however, investigate how logical compounding might affect truth-makers (and falsity-makers); nor does he consider infinitary truth-makers as mathematical objects alongside finitary proofs.

⁶ Discussion with Robert Kraut has been helpful here.

to forge any genuine connection with the traditional matters of logical consequence and formal proof. To the metaphysician, to be sure, these might sound like only a logician's concern—but the present author believes that truth-maker theory can advance only by affording an alternative account of truth, which could be substituted for the usual Tarskian account in the major branches of formal semantics and mathematical logic. Our truth-makers are intended to be the most economical way, logically and mathematically, of *reifying different realizations of truth-conditions* of logically complex sentences.⁷

On our account, it is also the case that any truth-maker Θ for a sentence φ is so constituted that the resulting truthmaking is an *internal* relation, in the sense of Armstrong (2004: section 1.3.3). As Armstrong puts it,

I mean by calling a relation internal that, given just the terms of the relation, the relation between them is necessitated. . . . I suggest it is an attractive ontological hypothesis that such a relation is no addition of being. Given just the terms, we are given the ontology of the situation.

That is what happens here: given just Θ and φ , we are given the ontology of the truth-making situation. The construction Θ makes φ true as a matter of necessity. “[G]iven just the terms of the [truth-making] relation, the relation between them is necessitated.” Justification for this claim (as for other claims made in this section) will have to await the laying out of the logical theory promised, which characterizes exactly what kind of object a truth-maker is.

When *they are finite* our truth-makers are, as it happens, exactly like proofs, since they possess all the properties that one requires of a proof: that it should admit of a mechanical check for correctness, and that one should be able to “read off” from it *what* it establishes, and *from* what. When, however, our truth-makers are perforce infinite—which can happen when there are infinitely many individuals in the domain of the interpretation in question⁸—what they reveal is that the grounds of truth (of the claim φ in question) are themselves infinite. This is why it would be a mistake to think of “grounds for truth,” as codified by a truth-maker, in an epistemic way. For a given claim φ , it can be the case that all its truth-makers are infinite even if φ can be furnished with a finitary proof within some axiomatic theory of the subject-matter, in which all the axioms (drawn from some decidable set Δ) strike the theorist as self-evident. Many a logically complex claim A , whose grounds of truth are essentially infinite, can qualify as self-evident and be taken as an axiom, i.e. as a member of Δ . In such a case, the proof-based epistemological route to the claim φ in question, by taking A as a deductive starting point, begins with a tremendously abbreviating short cut.⁹ Proofs are not always

⁷ Comments from Fraser MacBride have been helpful in prompting this description, for the metaphysician, of the aims of the brand of truth-making on offer here.

⁸ In some cases, a truth-maker can be finite even when the domain is infinite. Such can be the case with, say, an existential claim. But in an infinite domain any universal claim (or negative existential) can have only infinite truth-makers. Details will emerge below.

⁹ Two examples: *0 is not the successor of any natural number*; and *every natural number has a successor*. For the vast majority of mathematicians, these are self-evident, and are adopted as axioms. That is to

truth-makers, in the sense of “truth-maker” to be essayed here. And truth-makers are not always proofs.

An account of truth-makers must mesh with one’s accounts in two other areas of philosophical inquiry—the theory of (perceptual) knowledge, and the logician’s theory of logical consequence. The approaches already cited have something substantial to contribute to the former, but nothing, so far as we can see, to the latter. Our approach should be hospitable to including the best of what is on offer concerning the former. It seeks also to make an innovative contribution to the latter. In the section titled “Conclusion: open questions” of his Introduction to Lowe and Rami (2009), Rami does not list as an open problem that of making any kind of connection between *truth-making*, on the one hand, and, on the other hand, *preservation of truth from premises to conclusions of valid arguments*. From our perspective on the problem of truth-makers, however, this connection is key, as the reader will see in section 14.8.

In developing the beginnings of a theory of truth-makers as structured, logically complex entities, we should remind ourselves of the need for some methodological caution. One should not be over-ambitious. It took an impressive cast of logicians from Frege through to Tarski and Gentzen to master the details of first-order languages involving \neg , \wedge , \vee , \rightarrow , \exists , \forall , and $=$ on both the semantic and the deductive sides. Adequate formal semantics for the modal operators (involving possible worlds) had to await the work of Kripke in the 1960s. And although there are yet more recent formal semantic accounts to be had for such expressions as the counterfactual conditional, the relevant conditional, verbs of propositional attitude, and temporal operators, there is by no means a consensus on how such expressions are to be treated, either in formal semantics or in proof theory. So we restrict ourselves at the outset to a pursuit of an account of truth-makers that will at least work for the logician’s standard extensional, first-order language. The matter to be judged is whether the account succeeds *there*—not on whether, as presented here, it can handle (say) the alethic modal operators, or verbs of propositional attitude, or tense operators.

14.5 SOME BASIC NOTIONS

All the notions to be defined here can be defined much more rigorously, in a way that would meet the requirements of even the most captious mathematical logician. We are

say, they serve as their own *epistemic* grounds. Yet their grounds *of truth* involve each and every natural number, and hence are infinite.

One can seek to mitigate the gap between epistemic grounds and grounds of truth by seeking even deeper foundations, as the logicist does, for the mathematical axioms that are usually taken as starting points for mathematical proofs. Those axioms can then be derived in a non-trivial fashion as theorems, from yet more “obvious” and logical-looking first principles in the logicist’s theory. (See e.g. the constructive logicist derivation of the Peano postulates provided by Tennant (1987) and Tennant (2008).) Still, however, the grounds of truth of the latter principles are in all likelihood going to be infinite.

opting in the main part of this study for a slightly more informal presentation, by way of judiciously chosen examples rather than by completely general formal definitions, in order to ease the flow of philosophical discussion. (Section 14.10 provides a much more rigorous formal definition of truth-makers and falsity-makers.)

14.5.1 Saturated terms and formulae

In general, terms and formulae of the object-language may contain free variables. If they do, then they are called *open*. Those that are not open are called *closed*. A closed formula is called a *sentence*. A closed term may be called a (simple or complex) *name*. The semantic value of a name, when it has one, is an *individual*, which the name is then said to *denote*. The semantic value of a sentence, when it has one, is a *truth-value*, and the sentence is said to be *true* or *false* according as that value is *T* or *F*.

Closing an open term or formula involves substituting closed terms (of the object-language) for free occurrences of variables. Thus one could substitute the object-linguistic name j for the free occurrence of the variable x in the open term $f(x)$, to obtain the closed term $f(j)$ (“the father of John”). Or, to complicate the example slightly, one could substitute the closed object-linguistic term $m(j)$ for that free occurrence, to obtain the closed term $f(m(j))$ (“John’s maternal grandfather”). Likewise, an open formula, say $L(x, y)$, may be closed by substituting closed terms for its free occurrences of variables. One such closing would result in the sentence $L(m(f(j)), f(j))$ (“John’s paternal grandmother loves his father”).

Here we shall introduce an operation on open terms and formulae analogous to the operation of closing, but importantly different from it. The new operation will be called *saturation*. Like the operation of closing, the operation of saturation gets rid of all free occurrences of variables within an object-linguistic term or formula. But the way it does so is importantly different. Instead of substituting closed object-linguistic terms for free occurrences of variables, saturation is effected by substituting *individuals from the domain* for those free occurrences. Thus if α and β are individuals from the domain, one saturation of the open formula $L(x, y)$ would be $L(\alpha, \beta)$. Another one would be $L(m(f(\alpha)), f(\alpha))$, where the saturation is effected by substituting the saturated terms $m(f(\alpha))$ and $f(\alpha)$ for the free occurrences of the variables x and y respectively.

When the domain D supplies all the individuals involved in a saturation operation, the resulting saturated terms are called *saturated D -terms*, and the resulting saturated formulae are called *saturated D -formulae*.

Saturated terms and formulae are object-linguistic and metalinguistic hybrids. But, as mathematical objects, they are well-defined. When one treats, in standard Tarskian semantics, of assignments of individuals to variables, one is assuming such well-defined status for ordered pairs of the form $\langle x, \alpha \rangle$, where x is a free variable of the

object-language, and α is an individual from the domain of discourse.¹⁰ Since standard semantics is already committed to the use of such hybrid entities, it may as well take advantage of similar hybrid entities such as saturated terms and formulae.

We shall be taking advantage of them by having them feature in the rules of inference on the right in our description of an illustrative model below. Indeed, such rules will form a constitutive part of the model in question, as will emerge in due course.

14.5.2 Verification and falsification of sentences in models

We shall first describe, in general terms, *interpretations*, or *models*, of a first-order language.

A *model* consists of

- (i) a domain of individuals;
- (ii) a denotation mapping for names (if there are any names in the object-language);
- (iii) the structure that consists in primitive predicate-extensions; and
- (iv) the structure that consists in the mappings represented by primitive function-signs.

The names, predicates, and function-signs make up the *extra-logical vocabulary* that is being interpreted by the model in question.

For purposes of illustration, we shall confine ourselves to one-place predicates. So our expressive resources will be quite modest.

We shall now give a simple example of a model (see the figure on the following page). To the left will be a diagram, which can be thought of as the model M itself. The large dots will be the individuals; each one-place predicate extension will be represented by a box, as is usual with Euler–Venn diagrams. (If function-symbols were involved, then each one-place mapping would be represented by arrows—a different style of arrow for each mapping.)

So, for our toy example, we first choose a domain of individuals (here, three). They are labeled α , β , and γ in the *metalanguage*. The extensions of the two monadic predicates F and G are indicated by the boxes in the diagram. That diagram on the left, labeled M , “is” the model.

To the right of the diagram we provide an alternative presentation of the very same model. It takes an “inferentialist” form, using saturated atomic formulae (see section 14.5.1). In the top row to the right of the diagram are three “ M -relative” rules of inference, whose premises are *saturated identity-formulae* (see section 14.5.1 again) and whose conclusions are \perp . These rules ensure that the individuals are pairwise distinct.

¹⁰ This is true not only of Tarski’s original treatment (Tarski 1956; first published, in Polish, in 1933), which invoked *infinite sequences* of individuals correlated with object-linguistic variables, but also of the treatment (in Tennant 1978) of Tarski’s approach that appeals, more modestly, to *finitary* assignments of individuals to the free variables in a formula.

In the next two rows to the right of the diagram are some rules that specify the M -extensions of two predicates F and G . These rules can be “read off” from the diagram on the left. Note that this list of rules is exhaustive, dealing with all possible cases, both positive and negative, generated by combining a monadic predicate letter with an individual from the domain.¹¹

$ \begin{array}{ c } \hline M \\ \hline \begin{array}{ c c c } \hline \begin{array}{c} \bullet \\ \alpha \end{array} & \begin{array}{c} \bullet \\ \beta \end{array} & \begin{array}{c} \bullet \\ \gamma \end{array} \\ \hline \end{array} \end{array} $	$ \frac{\alpha = \beta}{\perp}_M $	$ \frac{\alpha = \gamma}{\perp}_M $	$ \frac{\beta = \gamma}{\perp}_M $
	$ \frac{}{F\alpha}_M $	$ \frac{F\beta}{\perp}_M $	$ \frac{F\gamma}{\perp}_M $
	$ \frac{}{G\alpha}_M $	$ \frac{}{G\beta}_M $	$ \frac{G\gamma}{\perp}_M $

The rules that have saturated formulae as their conclusions are degenerate examples of *truth-makers*; while the rules that have \perp as conclusion are degenerate examples of *falsity-makers*. Just as truth depends on the existence of a truth-maker, so too does falsity depend on the existence of a falsity-maker.¹²

What about true negative claims? Do they fail to have truth-makers? As Rami states the problem (Lowe and Rami 2009: 15),

Most notoriously difficult are so-called negative propositions: basically, negative predications such as the proposition that grass is not black and negative existentials such as the proposition that there are no unicorns. *Intuitively these propositions do not have truth-making entities if they are true.* So it is a counterintuitive consequence if the truth-maker theorist is forced to find such truth-making entities. (Emphasis added—NT)

We reject the italicized claim, as well as the one following it. We have just seen a falsity-maker for the claim that (in the model M) the individual γ has the property G :

$$\frac{G\gamma}{\perp}_M$$

With one step of \neg -Verification (analogous to \neg -Introduction in natural deduction) this can be extended so as to become a *truth-maker* for the *negative* claim $\neg G\gamma$, as follows:

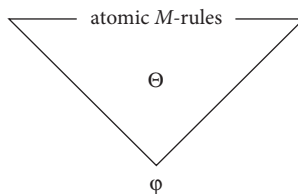
¹¹ This means that, if one had a *binary* predicate to interpret, and the same three individuals in the domain, one would have to furnish the right kind of rule for each of the *nine* ($= 3^2$) possible combinations of that predicate with *ordered pairs* of individuals from the domain.

¹² We prefer the term “falsity-maker” to the term “false-maker” of Armstrong (2004: section 1.4). In contrast with “truth” we need the noun “falsity” rather than the adjective “false.”

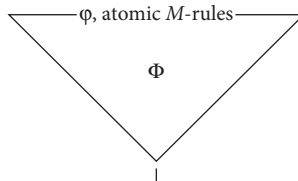
$$\frac{\frac{\frac{\text{---}}{\text{---}}^{(1)} G\gamma}{M} \perp}{\text{---}}^{(1)} \neg G\gamma$$

A striking feature of our theory of truth-makers is that it is really a theory, *simultaneously*, both of truth-makers and of falsity-makers. Ironically, earlier theorists of truth-making have focused their attention on the atomic case in the conviction that Tarskian recursion will take care of the logical operators, without our having to find truth-makers for logically complex sentences. In doing so, however, they have missed a golden opportunity to avail themselves of the *recursive interdependence* of truth-making and falsity-making.¹³ Ours is a unified theory of constructions that are either truth-makers or falsity-makers, as will emerge below.

For the time being, it will be helpful if the reader were to bear in mind that the overall structure of an M -relative truth-maker Θ for φ is that of a rule-governed passage *from* atomic information about (or, better: *from*) the model M to the sentence φ as conclusion:



and that the overall structure of an M -relative falsity-maker Φ for φ is that of a rule-governed passage *from* the sentence φ that it makes false, plus atomic information about (or from) the model M , to the conclusion \perp :



¹³ Throughout the recent revival of interest in truth-makers and truth-making, there has been a curious blindness to the possibility that truth-makers might be as variegated and as structurally complex as proofs are. In his Introduction to Lowe and Rami (2009), Rami writes (10),

Among truth-maker theorists there is considerable disagreement concerning the question which kinds of entity should be accepted as truth-*makers*: some only accept states of affairs as truth-making entities;[fn] some only accept individuals and states of affairs;[fn] and some only accept individuals and particular properties.[fn]

Notably absent from this list of extant possibilities is any conception of truth-makers along the lines proposed here.

So a truth-maker looks rather like a proof (but from very special premises), while a falsity-maker looks like a *reductio*, or disproof (but again, *modulo* very special premises). But it is to be borne in mind that each kind of construction can contain within it constructions of the other, complementary, kind. Truth-makers can be embedded within falsity-makers, and vice versa. The two kinds of construction are *recursively interdependent*, or *intercalated*.

Moreover, each kind of construction can be infinitary. The reader is reminded, once again, that this means that these constructions are intended to articulate the constitution of truth-making, rather than to serve as epistemic warrants.

In speaking of “rule-governed passages,” the rules to which we advert are what we shall call *evaluation rules*.¹⁴ They are set out in graphic form in section 14.7, and in section 14.10 they are set out again in the form of clauses in an inductive definition of the notions of truth-maker and of falsity-maker. These rules, especially in their graphic form, are similar, but not identical, to the familiar introduction and elimination rules of natural deduction.¹⁵ Corresponding to the introduction rules will be our rules-for-evaluating-as-true; and corresponding to the elimination rules will be our rules-for-evaluating-as-false. And the latter kind of evaluation will consist in “deriving” (in a possibly infinitary way) the conclusion \perp (absurdity) from the sentence that is to be evaluated as false.

14.6 EXAMPLES OF *M*-RELATIVE TRUTH-MAKERS

We shall immediately show what truth-makers look like (in the finite case), by giving easy examples before stating the general rules that allow for their construction.¹⁶ The model *M* in section 14.5.2 clearly makes the following sentences true:

$$\exists xFx \quad \forall x(Fx \rightarrow Gx)$$

In the model *M*, there is an *F*; and all *F*s are *G*s. Here is an *M*-relative truth-maker for the first of these claims (indeed, it is the *only* one available):

$$\frac{\frac{\text{---}^M}{F\alpha}}{\exists xFx}$$

¹⁴ Again, the reader is not to think of evaluation as an epistemic process, or as involving human judgment. It is meant here in a more neutral sense, as when a logician says that a universal quantification $\forall xG(x)$ over all the real numbers is to be “evaluated” as true if its predicate *G* is true of each real number. If preferred, the reader is welcome to read “evaluation” as “constitution” throughout (and similarly for cognates).

¹⁵ The first detailed and systematic presentation of these evaluation rules, with discussion of their relation to the rules of natural deduction, can be found in Tennant (2010). A reasonably well-cultured germ of the main ideas is to be found in Tennant (1997: section 11.3.2).

¹⁶ The finitude of our example must not lure the reader into thinking that truth-makers in general are like epistemic warrants. We have been at pains to distinguish the two notions. That having been said, however, it is the case that *finitary* truth-makers can serve as epistemic warrants. This point was made earlier.

And here is a truth-maker for the second:

$$\frac{
 \frac{\frac{}{M} G\alpha}{F\alpha \rightarrow G\alpha} \quad
 \frac{\frac{}{M} G\beta}{F\beta \rightarrow G\beta} \quad
 \frac{\frac{\frac{}{(1)} F\gamma}{\perp} M}{F\gamma \rightarrow G\gamma} (1)
 }{\forall x(Fx \rightarrow Gx)} D = \{\alpha, \beta, \gamma\}$$

The first truth-maker is self-explanatory. The second deserves some comment. It is not the only M -relative truth-maker for $\forall x(Fx \rightarrow Gx)$, since one can provide a different truth-maker for the middle instance $F\beta \rightarrow G\beta$:

$$\frac{\frac{\frac{}{(1)} F\beta}{\perp} M}{F\beta \rightarrow G\beta} (1)$$

The truth of the saturated material conditional $F\beta \rightarrow G\beta$ is, as it were, overdetermined in the model M , since its antecedent is false and its consequent is true. We chose to proceed via the truth of the consequent. The final step of inference in our truth-maker, to the conclusion $\forall x(Fx \rightarrow Gx)$, was taken from *all the instances* of the latter in the domain D of the model M . Those instances are the three saturated formulae

$$F\alpha \rightarrow G\alpha, \quad F\beta \rightarrow G\beta, \quad F\gamma \rightarrow G\gamma$$

That is how we give expression to the fact that α , β , and γ are *all the individuals there are* in M : we allow (the evaluation-analogue of) “universal introduction” *from* all the instances of the claim *to* the universal claim in question.¹⁷

14.7 EVALUATION RULES

When applying any evaluation rule whose conclusion is \perp , the major premise is to *stand proud*, with no “evaluation work” above it.¹⁸ This ensures that all truth-makers

¹⁷ Likewise, we allow (the evaluation-analogue of) “existential elimination” from the major (existential) premise, plus falsity-makers for all instances thereof, to the conclusion \perp . See section 14.7.3.

¹⁸ Again, “work” is here not to be construed as the result of intellectual effort, or an outcome representing some sort of epistemic achievement. It refers only to possible constructions (truth-makers) that constitute the grounds for truth of the sentence in question, in the ontological or metaphysical sense explained earlier.

(i.e. evaluations-as-true) and all falsity-makers (i.e. evaluations-as-false) are in *normal form*, in the proof-theorist's sense.¹⁹

Another requirement on applications of these rules is that there may not be any “vacuous discharges” of assumptions. If a rule indicates that an assumption of a certain form may be discharged, then, when the rule is applied, just such an assumption must already be in place (i.e. it must already have been used) and hence be available to be discharged.²⁰

14.7.1 Rules for primitive saturated formulae

$$\frac{A(\alpha_1, \dots, \alpha_n) \quad t_1 = \alpha_1 \quad \dots \quad t_n = \alpha_n}{A(t_1, \dots, t_n)}$$

where A is a primitive n -place predicate, t_1, \dots, t_n are saturated D -terms, and $\alpha_1, \dots, \alpha_n$ are individuals in the domain.

$$\frac{\frac{\frac{}{A(\alpha_1, \dots, \alpha_n)} (i)}{\perp} M \quad t_1 = \alpha_1 \quad \dots \quad t_n = \alpha_n \quad A(t_1, \dots, t_n)}{\perp} (i)$$

where A is a primitive n -place predicate, t_1, \dots, t_n are saturated D -terms, and $\alpha_1, \dots, \alpha_n$ are individuals in the domain.

¹⁹ This way of ensuring normal form—but of ordinary *deductive proofs*—was first treated in Tennant (1992). Here we are applying the idea to the kinds of construction that we are calling truth-makers and falsity-makers.

²⁰ This is a very important precondition for the *relevance* of premises to conclusions in natural deduction. It was stated in Tennant (1979). It is exactly the proof-theoretic insight that is needed in order to put appropriate proof-theoretic resources at the disposal of the truth-maker theorist. For the operation of determining the truth-value of a compound from the truth-values of its immediate constituents (i.e. subsentences, or instances of quantifications) is, crucially, a *relevant* one. The truth-value of the compound *depends on* those of its immediate constituents. No “proof-like” conception of truth-makers (and falsity-makers) could possibly work if one did not pay attention to this fact, and *build that very feature into* one's account.

None of this, of course, is to be construed as turning truth-makers (or falsity-makers) into epistemic warrants or justifications. They remain *metaphysical* objects. They are constituted by the mathematics of finite and infinite labeled trees, whose labels (which are often saturated formulae) can embed individuals. Proof-theoretic notions happen to be applicable to these trees only because such trees have been studied extensively in proof theory. But, whereas the proofs of traditional proof theory have always been understood as epistemic warrants, the more generalized trees that we are dealing with here, and which we are calling truth-makers and falsity-makers, are being put into the service of metaphysics and ontology. The truth-making and falsity-making involved can be as “epistemically transcendent” as any realist reader might wish. (This anti-realist author is really running with the wolves here.)

$$\frac{f(\alpha_1, \dots, \alpha_n) = \alpha \quad u_1 = \alpha_1 \quad \dots \quad u_n = \alpha_n}{f(u_1, \dots, u_n) = \alpha}$$

where f is a primitive n -place function sign, u_1, \dots, u_n are saturated D -terms and $\alpha, \alpha_1, \dots, \alpha_n$ are individuals in the domain

$$\frac{\frac{\text{---}^{(i)}}{f(\alpha_1, \dots, \alpha_n) = \alpha} \quad M}{\perp} \quad \frac{u_1 = \alpha_1 \quad \dots \quad u_n = \alpha_n \quad f(u_1, \dots, u_n) = \alpha}{\perp}^{(i)}$$

where f is a primitive n -place function sign, u_1, \dots, u_n are saturated D -terms, and $\alpha, \alpha_1, \dots, \alpha_n$ are individuals in the domain.

$$\frac{\text{---}}{\alpha = \alpha} \text{ where } \alpha \text{ is an individual in the domain}$$

14.7.2 Rules for saturated formulae with a connective dominant

The reader will recognize these rules as simply codifying the left-to-right readings of the rows of the respective truth-tables. Remember: all major premises for eliminations stand proud; and vacuous discharges of assumptions are prohibited.

$$\frac{\frac{\text{---}^{(i)}}{\varphi} \quad \vdots \quad \perp}{\neg\varphi}^{(i)} \quad \frac{\neg\varphi \quad \varphi}{\perp}$$

$$\frac{\varphi \quad \psi}{\varphi \wedge \psi} \quad \frac{\frac{\text{---}^{(i)}}{\varphi} \quad \vdots \quad \perp}{\varphi \wedge \psi}^{(i)} \quad \frac{\frac{\text{---}^{(i)}}{\psi} \quad \vdots \quad \perp}{\varphi \wedge \psi}^{(i)}$$

$\frac{\varphi}{\varphi \vee \varphi} \quad \frac{\psi}{\varphi \vee \psi}$	$\frac{\frac{\frac{\vdots}{\varphi \vee \psi} \perp}{\vdots} \perp \quad \frac{\frac{\vdots}{\psi} \perp}{\vdots} \perp}{\perp}^{(i)}$
---	--

$\frac{\frac{\vdots}{\psi} \perp}{\varphi \rightarrow \psi} \quad \frac{\frac{\vdots}{\perp}}{\varphi \rightarrow \psi}$	$\frac{\frac{\frac{\vdots}{\varphi \rightarrow \psi} \perp}{\vdots} \perp \quad \frac{\frac{\vdots}{\varphi} \perp}{\vdots} \perp}{\perp}^{(i)}$
--	--

14.7.3 Rules for saturated formulae with a quantifier dominant

The reader will recognize these rules as capturing the two Wittgensteinian insights often conveyed to students when explaining how the two main quantifiers are to be understood: universal quantifications behave like conjunctions of all their instances; and existential quantifications behave like disjunctions of the same. Once again: all major premises for eliminations stand proud; and vacuous discharges of assumptions are prohibited.

$\frac{\psi(\alpha_1) \dots \psi(\alpha_n) \dots}{\forall x \psi(x)}$	$\frac{\frac{\frac{\vdots}{\psi(\alpha)} \perp}{\vdots} \perp}{\forall x \psi(x)}^{(i)}$
---	--

where $\alpha_1, \dots, \alpha_n \dots$ are all the individuals in the domain

$\frac{\vdots}{\psi(\alpha)}$	$\frac{\frac{\frac{\vdots}{\psi(\alpha_1)} \perp}{\vdots} \perp \quad \dots \quad \frac{\frac{\vdots}{\psi(\alpha_n)} \perp}{\vdots} \perp}{\perp}^{(i)}$
-------------------------------	---

where $\alpha_1, \dots, \alpha_n \dots$ are all the individuals in the domain

14.7.4 Comments on the rules and the constructions they constitute

One can quickly incur sideways spread in writing down a detailed truth-maker or falsity-maker. This feature militates against the actual construction of these otherwise very illuminating and detailed constructions for sentences (and saturated formulae) relative to a given model M . As soon as one has three or more individuals in the domain of M , along with nested quantifiers (especially when they occasion the use of the two rules that require investigation of all instances of a quantified claim), the blow-up, in the form of sideways spread, is prohibitive. But the resulting construction is only ever as *deep* as the longest branch within the analysis tree of the sentence (or saturated formula) being evaluated.

Moreover, in cases where the domain is *infinite*, some of these truth-makers and falsity-makers will contain steps (for the truth-maker for a universal, or the falsity-maker for an existential) that require *infinitely many* subordinate constructions (in the form of truth-makers or falsity-makers for instances of the quantified claims in question). In such cases the constructions cannot be written down. Instead, they exist only as infinitary mathematical objects: labeled trees (where the labels are at least finite!) that can have infinite branching, albeit only with branches of finite length.²¹ Ultimately, the present “inferentialist” approach to formal semantics via the truth-makers and falsity-makers illustrated above requires no more powerful mathematical machinery than is needed in order to vouchsafe the existence of these (rather modest) kinds of infinitary object.

The combinatorial mathematics that is required in order to vouchsafe our truth-makers and falsity-makers is of very modest consistency strength. In any possible world containing both the individuals involved in the model in question and the symbols that are interpreted by the model, the truth-makers and falsity-makers themselves can be taken to form part of one’s ontology at no further cost. To the extent that the “mathematics of trees” (like the arithmetic of the natural numbers) is part of any reasonably sophisticated conceptual scheme, one can rely on our constructions existing whenever their ultimate constituents exist.

14.7.5 Reflexive stability

Any account of truth-makers faces the potential problem of reflexive instability. If Θ is held, on the account in question, to be a truth-maker for φ , then the question arises: What would be the truth-maker, if any, for the proposition that Θ is indeed a

²¹ Note that the branches, all of them finite, are bounded in length by the length of the longest branch in the analysis tree of the sentence being evaluated.

truth-maker for φ ? In facing this question, we are fortunately on firm ground. Our definition of the notions of (M -relative) *truth-maker* and of (M -relative) *falsity-maker* is a co-inductive one. (See section 14.10.) Hence the metalinguistic conclusion “ Θ is an M -relative truth-maker for φ ” (abbreviated as $\mathcal{V}_M(\Theta, \varphi)$) can be displayed as the conclusion of a *metalinguistic truth-maker*, which methodically unravels, step by step (and in the appropriate partial order—even if it is an “infinitely sideways-branching” one), the inductive-definitional grounds for the truth of the theoretical claim $\mathcal{V}_M(\Theta, \varphi)$ itself! This is analogous to the way in which one can *prove*, in a weak metatheory, that a (here, finite) tree-like object-linguistic construction is a *proof* (in Peano arithmetic, say) of a particular theorem of first-order arithmetic. It is also analogous to the way in which, in a simple language of color-predication, one can provide a truth-maker for the claim that a particular object α , which happens to be red, is colored. The language contains the analytic, atomic inferential rule

$$\frac{t \text{ is red}}{t \text{ is colored}}$$

If in the model M the individual α is red, then we have recourse to the M -relative atomic rule of inference

$$\frac{\text{—————}^M}{\alpha \text{ is red}}$$

This permits the formation of the M -relative truth-maker

$$\frac{\frac{\text{—————}^M}{\alpha \text{ is red}}}{\alpha \text{ is colored}}$$

for the claim that α is colored. In analogous fashion, within a metatheory containing logically atomic inductive clauses governing the formation of truth-makers and falsity-makers, we can build up truth-makers for metalinguistic claims of the form $\mathcal{V}_M(\Theta, \varphi)$. So reflexive stability is ensured.

14.8 LOGICAL CONSEQUENCE DEFINED IN TERMS OF TRUTH-MAKERS

The classical conception of logical consequence, due to Tarski (1956 [1936]), requires only *preservation of truth*.

Definition 1 (Tarski consequence)

ψ is a logical consequence of Δ

if and only if

for every interpretation M , if every member of Δ is true in M , then ψ is true in M also.

It is worth noting that the foregoing definiens (the right-hand side of the biconditional) is often written as

for every interpretation M , if M makes every member of Δ true, then M makes ψ true also.

This invites the reader to (mis)construe the model M itself as a truth-maker. But that would be a serious mistake, resulting from taking the gloss “ M makes ϕ true” much too literally. The model M is too gross an object to serve as a genuine (even M -relative!) truth-maker. Moreover, on this (mis)construal, M would be serving as *the* truth-maker for each and every sentence ϕ that is true in M . But the whole point of a more refined theory of truth-makers is to be able to distinguish (with respect to some same model M , whatever it may be) the *different* (but still M -relative) truth-makers that can be had by one and the same sentence ϕ , and, of course, also to distinguish among the different (but still M -relative) truth-makers that can be had by different sentences. Thinking of M as the sole M -relative truth-maker would be to obliterate all such possible refinements. In the special case where M were the actual world, it would be like taking the Great Fact as the only truth-maker for each and every truth.

The foregoing model-theoretic definition of logical consequence can be stated more formally as follows (where $M \models \phi$ symbolizes “ ϕ is true in M ”):

$$\begin{aligned} \Delta \models \psi \\ \Leftrightarrow \\ \forall M ([\forall \phi \in \Delta \ M \models \phi] \Rightarrow M \models \psi) \end{aligned}$$

This definition yields no insight, however, into how the grounds of truth-in- M of the conclusion ψ might be related to, or determinable from, the grounds of truth-in- M of the premises in Δ (relative to any of the models M interpreting the extralogical vocabulary involved).²²

There is an alternative conception of logical consequence, fashioned for intuitionistic logic (and for finite sets of premises), which sought to address what was taken to be this deficit of insight, and which is owed to Prawitz (1974).

²² The word “determinable” here is to be taken in a constitutive, not epistemic, sense. One should certainly not expect the method of determination, in general, to be effective.

Definition 2 (Prawitz consequence)

ψ is a logical consequence of $\varphi_1, \dots, \varphi_n$

if and only if

there is an effective method f such that for every interpretation M , and for all M -warrants π_1, \dots, π_n for $\varphi_1, \dots, \varphi_n$ respectively, $f(\pi_1, \dots, \pi_n)$ is an M -warrant for ψ .

More formally:

$$\varphi_1, \dots, \varphi_n \vdash_P \psi$$

$$\Leftrightarrow$$

$$\exists \text{eff}. f \forall M \forall \pi_1 \dots \forall \pi_n [((\mathcal{P}_M(\pi_1, \varphi_1) \wedge \dots \wedge \mathcal{P}_M(\pi_n, \varphi_n)) \Rightarrow \mathcal{P}_M(f(\pi_1, \dots, \pi_n), \psi))]$$

Here, $\mathcal{P}_M(\pi, \varphi)$ means that the construction π is an M -warrant for the sentence φ . On Prawitz's account, M is taken to be a so-called *basis* consisting of atomic rules of inference, which together provide some measure of interpretation of the primitive extra-logical expressions involved in them. Prawitz's conception of warrant is anti-realist in inspiration. Warrants are the canonical proofs that Dummett marshalls in arguments that are intended to motivate the adoption of anti-realist semantics. Dummett argues that one should regard as licit only the *recognizable* obtaining of the conditions of truth of a sentence—such as is displayed by a canonical proof. And, given this constraint on manifestation of understanding, he argues further, the Principle of Bivalence is infirmed.²³

So it is important to realize that Prawitz is not in the business of distinguishing metaphysical or ontological grounds for truth (in the constitutive sense essayed here) from justificatory grounds. His warrants are epistemic objects. And for Prawitz, as for Dummett, the truth of any claim *consists in* its possessing such a warrant. As an anti-realist, Prawitz makes no distinction between the constitutive grounds that a realist might believe in, and the contrasting epistemic grounds that are the only grounds the anti-realist can recognize as obtaining. But that should not deter us from trying to appreciate an important feature of Prawitz's characterization of logical consequence, and from trying to apply it profitably in our own characterization of the role that truth-makers can be made to play in a notion of logical consequence more congenial to the realist.

On Prawitz's definition of logical consequence, one is able to interpret the normalization theorem for intuitionistic natural deductions as furnishing a proof of the *soundness* of that proof-system. For the normalization procedure (call it ν) itself qualifies as the

²³ For a critique of this "manifestation argument," and an attempt to improve upon it, see Tennant (1997: ch. 7).

effective method f called for on the right-hand side of his definition. To see this, suppose one has an intuitionistic natural deduction Π of the conclusion ψ from (possibly logically complex) premises $\varphi_1, \dots, \varphi_n$. We need to convince ourselves that Π will preserve warranted assertibility from its premises to its conclusion. So suppose further that one furnishes M -warrants π_1, \dots, π_n (which take the form of closed canonical proofs using atomic rules in M) for $\varphi_1, \dots, \varphi_n$ respectively. Append each warrant to the assumption-occurrences of its conclusion in Π :

$$\begin{array}{c} \pi_1 \quad \dots \quad \pi_n \\ \varphi_1 \quad \dots \quad \varphi_n \\ \hline \Pi \\ \psi \end{array}$$

Then by normalizing the result one obtains an M -warrant π for the overall conclusion ψ :

$$\begin{array}{c} \pi_1 \quad \dots \quad \pi_n \\ \varphi_1 \quad \dots \quad \varphi_n \\ \hline \Pi \\ \psi \end{array} \quad \text{normalizes to} \quad \begin{array}{c} \pi \\ \psi \end{array}, \quad \text{where } \pi = v(\pi_1, \dots, \pi_n, \Pi)$$

What the foregoing reasoning shows is that

$$\varphi_1, \dots, \varphi_n \vdash_I \psi \Rightarrow \varphi_1, \dots, \varphi_n \models_P \psi.$$

This is the promised soundness result for intuitionistic deduction, with respect to Prawitz's notion of logical consequence.

The converse of this result:

$$\varphi_1, \dots, \varphi_n \models_P \psi \Rightarrow \varphi_1, \dots, \varphi_n \vdash_I \psi$$

remains as Prawitz's (as yet unproved) *completeness conjecture* for intuitionistic logic.

A very desirable feature of Prawitz's definition of consequence is how it *Skolemizes* (as $v(\pi_1, \dots, \pi_n, \Pi)$) that which is responsible for the (M -relative) truth of the conclusion ψ . On the classical conception, by contrast, one is given to understand only that if the premises are true, then the conclusion ψ is true also—but without giving one any idea as to how any (M -relative) truth-makers for ψ might depend on (M -relative) truth-makers for the premises.

The desirable feature of Prawitz's definition can, however, be taken over by the truth-maker theorist even if in the service of a more realist rather than anti-realist conception of truth and consequence. The truth-maker theorist can define logical consequence as follows.

Definition 3 (Truth-maker consequence)

ψ is a logical consequence of $\varphi_1, \dots, \varphi_n$

if and only if

there is a “quasi-effective” method f such that for every interpretation M , and for all M -relative truth-makers π_1, \dots, π_n for $\varphi_1, \dots, \varphi_n$ respectively, $f(\pi_1, \dots, \pi_n)$ is an M -relative truth-maker for ψ .

More formally:

$$\varphi_1, \dots, \varphi_n \models_T \psi$$

$$\Leftrightarrow$$

$$\exists q.\text{-eff.} f \forall M \forall \pi_1 \dots \forall \pi_n [((\mathcal{V}_M(\pi_1, \varphi_1) \wedge \dots \wedge \mathcal{V}_M(\pi_n, \varphi_n)) \Rightarrow \mathcal{V}_M(f(\pi_1, \dots, \pi_n), \psi)]$$

Here, $\mathcal{V}_M(\pi, \psi)$ means that the construction π is an M -relative truth-maker for the sentence ψ .

We have written “quasi-effective method f ” because of problems posed by the infinite case. In the finite case, the method will certainly be effective. But in the infinite case we have to countenance definable operations on infinite sets of constructions, and these by definition cannot be effective (in the sense in which Church’s Thesis states that all effective functions are recursive functions). The functions or methods that we shall employ, however, will be “effective-looking” in the infinite case, because of the obvious, orderly way in which one “follows the recipes” for truth-maker transformation within the infinite context. They are recipes obtained by smooth extrapolation from the finite case. In section 14.9 we give a simple finite example of how this truth-maker transformation is effected, by an appropriate analogue of the normalization procedure for proofs. It is an interesting research problem to characterize as precisely as possible the class of functions that we are here calling, informally, the “quasi-effective” ones.

Note that our statement of “truth-maker transformation” by valid argument is subtler than the so-called *entailment principle* (Lowe and Rami 2009: 26):

For every x , y and z : if x is a truth-maker for y and y entails z , then x is a truth-maker for z .

We do not subscribe to this principle. It is too sweeping. For example, on our account, the truth-maker for a true conjunction $\varphi \wedge \psi$ is a construction that contains as sub-constructions a truth-maker for φ and a truth-maker for ψ . But that does not make the truth-maker for the conjunction a truth-maker for either of its conjuncts. *Containing* a truth-maker for each conjunct does not amount to *being* a truth-maker for either of them.²⁴

²⁴ This puts us in disagreement with Restall (1996: 333), who writes, “There is also a desirable result connecting truthmaking and conjunction. If something makes both A and B true, then it also makes their conjunction true, and vice versa.”

14.9 AN EXAMPLE OF TRUTH-MAKER TRANSFORMATION BY VALID ARGUMENT

We saw earlier that the sentences

$$\exists xFx \quad \forall x(Fx \rightarrow Gx)$$

were both true in our chosen example of a very simple model M . Note that we have not yet laid out any M -relative truth-maker for $\exists xGx$. The reader is asked not to look back at the diagram for M , but to focus instead on the following deductive proof of the conclusion $\exists xGx$ from the above two sentences as premises. Note that this is a formal proof of the more familiar kind, in the system of natural deduction due to Gentzen (1934–35):

$$\frac{\frac{\frac{\forall x(Fx \rightarrow Gx)}{Fa \rightarrow Ga} \quad \frac{}{Fa}^{(1)}}{Ga}}{\frac{\exists xFx \quad \exists xGx^{(1)}}{\exists xGx}^{(1)}}$$

Consider now what happens when we supply the M -relative truth-makers constructed earlier for the two premises (or undischarged assumptions) of this deductive proof:

$$\frac{\frac{\frac{}{G\alpha}^M}{F\alpha \rightarrow G\alpha} \quad \frac{\frac{}{G\beta}^M}{F\beta \rightarrow G\beta} \quad \frac{\frac{\frac{}{F\gamma}^{(1)}}{\perp}^M}{F\gamma \rightarrow G\gamma}^{(1)}}{\forall x(Fx \rightarrow Gx)} \quad D = \{\alpha, \beta, \gamma\}$$

$$\frac{\frac{\frac{}{F\alpha}^M}{\exists xFx} \quad \frac{\frac{\frac{\forall x(Fx \rightarrow Gx)}{Fa \rightarrow Ga} \quad \frac{}{Fa}^{(1)}}{Ga}}{\exists xGx}^{(1)}}{\exists xGx}^{(1)}$$

When the truth-makers above are grafted so that their conclusions are superimposed upon the premise-occurrences of the same sentences in the formal proof below those truth-makers, we obtain a construction of a mixed kind—a combination of the truth-makers “above,” and the formal deductive proof “below”—that invites an analogue of the *normalization process* of the kind familiar to a proof-theorist.

The recipe for normalizing is much the same in the truth-maker setting as it is in the proof setting.²⁵ One treats applications of rules for evaluating-as-true (in M) as though they are applications of introduction rules, and similarly one treats applications of rules for evaluating-as-false (in M) as though they are applications of elimination rules. Only now, one has to countenance the possibility that one will be substituting *individuals* (from the domain of M) for free occurrences of variables in formulae, thereby forcing an erstwhile formal proof to involve *saturated formulae*.

In order to illustrate, we shall normalize the result of grafting the truth-maker for $\exists xFx$ and the truth-maker for $\forall x(Fx \rightarrow Gx)$ onto their respective undischarged assumption-occurrences in the formal proof “below” them. The resulting occurrences of $\exists xFx$ and $\forall x(Fx \rightarrow Gx)$ are then “maximal,” each standing as the conclusion of an “introduction” and as the major premise of the corresponding elimination. Dealing first with the maximal occurrence of $\exists xFx$, we apply the appropriately modified “reduction procedure” to obtain

$$\begin{array}{c}
 \frac{\frac{\frac{}{M} G\alpha}{F\alpha \rightarrow G\alpha} \quad \frac{\frac{}{M} G\beta}{F\beta \rightarrow G\beta} \quad \frac{\frac{\frac{}{(1)} F\gamma}{\perp} M}{F\gamma \rightarrow G\gamma}^{(1)}}{D = \{\alpha, \beta, \gamma\}} \\
 \frac{\frac{\forall x(Fx \rightarrow Gx)}{F\alpha \rightarrow G\alpha} \quad \frac{}{M} F\alpha}{G\alpha} \\
 \frac{}{\exists xGx}
 \end{array}$$

Note how the individual α has insinuated itself further into the resulting construction. We have left the realm of proofs, and entered a realm of constructions that contain actual individuals (like α) but are not (yet) actual truth-makers. For the construction at this stage is not yet in normal form.

Next we have to deal with the maximal occurrence of $\forall x(Fx \rightarrow Gx)$. We apply the reduction procedure for the universal quantifier in order to replace with an appropriate reduct the construction terminating on the lower occurrence of $F\alpha \rightarrow G\alpha$. The overall result is

$$\frac{\frac{\frac{}{M} G\alpha}{F\alpha \rightarrow G\alpha} \quad \frac{}{M} F\alpha}{G\alpha} \\
 \frac{}{\exists xGx}$$

²⁵ This remark holds even when we consider *infinitary* truth-makers “above,” in combination with (perforce finitary) proofs “below.” In the current illustration, however, we are dealing with finitary truth-makers, because M is finite.

in which the occurrence of $F\alpha \rightarrow G\alpha$ is now (newly) maximal. Applying the reduction procedure for the conditional, we obtain as our final result the (normal!) truth-maker

$$\frac{\text{---}^M}{G\alpha} \\ \hline \exists xGx$$

Now the reader is invited to look back at the model M . This is one of the truth-makers one could have given for the claim $\exists xGx$. The other possibility was

$$\frac{\text{---}^M}{G\beta} \\ \hline \exists xGx$$

But the truth-maker involving β was ignored by the normalization process. This is because the process was directed toward uncovering a truth-maker for $\exists xGx$ that was *implicit in* the combination of (i) the truth-makers actually furnished for $\exists xFx$ and for $\forall x(Fx \rightarrow Gx)$ —the first of which involved α , not β —and (ii) the deductive proof of $\exists xGx$ from those two claims as premises. The last-displayed truth-maker for $\exists xGx$ could not, of course, come to light as a result of this very focused process. It is the Skolemizing nature of the construction of grounds for the truth of the conclusion of a deductive proof, out of grounds for the truth of its premises, which underlies this ability to focus on what is relevant. We are now in a better position to appreciate how it is that a formal deduction enables one to see how the truth of a conclusion is contained in that of the premises—not, as Frege once wrote,²⁶ as beams are contained in a house, but rather as a plant is contained in the seed from which it grows. And we have helped explicate this insight for the benefit of the Fregean *realist*.

We are now in a position to reflect somewhat critically on the following claim:

Th[e] idea of a perfect parallelism of logical and ontological complexity is the misery of logical atomism . . . [W]e uphold the independence of ontological from logical complexity: ontologically complex objects (those having proper parts) are not for that reason also in some way logically complex, any more than there is reason to suppose that to every logically complex (true) sentence there corresponds an ontologically complex entity which makes it true. (Mulligan et al. 1984: 298)

With our account of truth-makers and falsity-makers as *appropriately structured* complex entities, we have sought to relieve this particular “misery” of logical atomism. We are not, of course, claiming that ontologically complex objects (those having proper parts) are also in some way logically complex. Rather, we are challenging the claim

²⁶ See Frege (1884; reprinted 1961: section 88).

that there is no “reason to suppose that to every logically complex (true) sentence there corresponds an ontologically complex entity which makes it true.” On the contrary, our theory of truth-makers shows that that is precisely the case.

14.10 A FORMAL CO-INDUCTIVE DEFINITION OF TRUTH-MAKERS AND FALSITY-MAKERS

Earlier we employed the formal expression $\mathcal{V}_M(\Pi, \varphi)$ to express the claim that Π is an M -relative truth-maker for the sentence (or saturated formula) φ . We shall now elevate the subscript M to be a separate argument of this predication, alongside its domain D of individuals. That way, M can be reconstrued as consisting of *saturated literals*, or the rules that codify them.

A *saturated literal* is a saturation of an atomic formula, or of the negation of an atomic formula. The model M will of course be both *complete* and *coherent*. The *completeness* of M (when it is construed as a set of rules) is a matter of containing, for each n -tuple $\vec{\alpha}$ of individuals drawn from the domain D of M , and for each n -place atomic predicate A , either the “positive” rule

$$\frac{}{A(\vec{\alpha})}^M$$

or the “negative” rule

$$\frac{A(\vec{\alpha})}{\perp}^M;$$

whereas the *coherence* of M is a matter of *not* containing, for any such tuple $\vec{\alpha}$ and predicate A , *both* of these rules. Our rule-theoretic conception of a model is therefore thoroughly Fregean and Tarskian, in assuming determinacy of all atomic facts. Indeed, we can consider M as ambiguous between its rule-theoretic characterization, on the one hand, and, on the other hand, the more usual Tarskian characterization, in terms of A -extensions that are subsets of the n -fold Cartesian product of the domain D of M .

Definition 4

The two notions $\mathcal{V}(\Pi, \varphi, M, D)$ and $\mathcal{F}(\Pi, \varphi, M, D)$ (respectively: “the construction Π is a truth-maker for the saturated formula φ modulo the domain D and the set M of saturated literals” and “the construction Π is a falsity-maker for the saturated formula φ modulo the domain D and the set M of saturated literals”) are co-inductively defined by the following metalinguistic axioms and rules of inference:

For saturated literals λ whose individuals are in D :

$$(\lambda\mathcal{V}) \frac{}{\mathcal{V}(\lambda, \lambda, \{\lambda\}, D)}$$

For positive saturated literals A whose individuals are in D :

$$(At\mathcal{F}) \frac{}{\mathcal{F}\left(\frac{\neg A \quad A}{\perp}, A, \{\neg A\}, D\right)} \quad \frac{}{\mathcal{F}\left(\frac{\neg A \quad A}{\perp}, \neg A, \{A\}, D\right)}$$

For non-atomic saturated formulae φ whose individuals are in D :

$$(\neg\mathcal{V}) \frac{\mathcal{F}(\Pi, \varphi, M, D)}{\mathcal{V}\left(\frac{\Pi}{\neg\varphi}, \neg\varphi, M, D\right)}$$

$$(\neg\mathcal{F}) \frac{\mathcal{V}(\Pi, \varphi, M, D)}{\mathcal{F}\left(\frac{\neg\varphi \quad \Pi}{\perp}, \neg\varphi, M, D\right)}$$

For saturated formulae φ_1, φ_2 of any degree of complexity (including positive saturated literals) whose individuals are in D :

$$(\wedge\mathcal{V}) \frac{\mathcal{V}(\Pi_1, \varphi_1, M_1, D) \quad \mathcal{V}(\Pi_2, \varphi_2, M_2, D)}{\mathcal{V}\left(\frac{\Pi_1 \quad \Pi_2}{\varphi_1 \wedge \varphi_2}, \varphi_1 \wedge \varphi_2, M_1 \cup M_2, D\right)}$$

$$(\wedge\mathcal{F}) \frac{\mathcal{F}(\Pi, \varphi_i, M, D)}{\mathcal{F}\left(\frac{\varphi_1 \wedge \varphi_2 \quad \Pi}{\perp}, \varphi_1 \wedge \varphi_2, M, D\right)} \quad (i = 1, 2)$$

$$(\vee\mathcal{V}) \frac{\mathcal{V}(\Pi, \varphi_i, M, D)}{\mathcal{V}\left(\frac{\Pi}{\varphi_1 \vee \varphi_2}, \varphi_1 \vee \varphi_2, M, D\right)} \quad (i = 1, 2)$$

$$(\vee\mathcal{F}) \frac{\mathcal{F}(\Pi_1, \varphi_1, M_1, D) \quad \mathcal{F}(\Pi_2, \varphi_2, M_2, D)}{\mathcal{F}\left(\frac{\varphi_1 \vee \varphi_2 \quad \Pi_1 \quad \Pi_2}{\perp}, \varphi_1 \vee \varphi_2, M_1 \cup M_2, D\right)}$$

$$(\rightarrow\mathcal{V}) \frac{\mathcal{V}(\Pi, \varphi_2, M, D)}{\mathcal{V}\left(\frac{\Pi}{\varphi_1 \rightarrow \varphi_2}, \varphi_1 \rightarrow \varphi_2, M, D\right)} \quad \frac{\mathcal{F}(\Pi, \varphi_1, M, D)}{\mathcal{V}\left(\frac{\Pi}{\varphi_1 \rightarrow \varphi_2}, \varphi_1 \rightarrow \varphi_2, M, D\right)}$$

$$(\rightarrow \mathcal{F}) \frac{\mathcal{V}(\Pi_1, \varphi_1, M_1, D) \quad \mathcal{F}(\Pi_2, \varphi_2, M_2, D)}{\mathcal{F}\left(\frac{\varphi_1 \rightarrow \varphi_2 \quad \Pi_1 \quad \Pi_2}{\perp}, \varphi_1 \rightarrow \varphi_2, M_1 \cup M_2, D\right)}$$

For ψ any unary saturated formula whose individuals are in D :

$$\begin{aligned} (\exists \mathcal{V}) & \frac{\mathcal{V}(\Pi, \psi_\alpha^x, M, D)}{\mathcal{V}\left(\frac{\Pi}{\exists x \psi}, \exists x \psi, M, D\right)} \quad \text{where } \alpha \in D \\ (\exists \mathcal{F}) & \frac{\{\mathcal{F}(\Pi_\alpha, \psi_\alpha^x, M_\alpha, D)\}_{\alpha \in D}}{\mathcal{F}\left(\frac{\exists x \psi \quad \{\Pi_\alpha\}_{\alpha \in D}}{\perp}, \exists x \psi, \bigcup_{\alpha \in D} M_\alpha, D\right)} \\ (\forall \mathcal{V}) & \frac{\{\mathcal{V}(\Pi_\alpha, \psi_\alpha^x, M_\alpha, D)\}_{\alpha \in D}}{\mathcal{V}\left(\frac{\{\Pi_\alpha\}_{\alpha \in D}}{\forall x \psi}, \forall x \psi, \bigcup_{\alpha \in D} M_\alpha, D\right)} \\ (\forall \mathcal{F}) & \frac{\mathcal{F}(\Pi, \psi_\alpha^x, M, D)}{\mathcal{F}\left(\frac{\forall x \psi \quad \Pi}{\perp}, \forall x \psi, M, D\right)} \quad \text{where } \alpha \in D \end{aligned}$$

The foregoing rules are the basis rules and the inductive rules in the co-inductive definition of the notions $\mathcal{F}(\Pi, \varphi, M, D)$ and $\mathcal{V}(\Pi, \varphi, M, D)$. To complete the definition, we have the *Closure clause*.

If $\mathcal{V}(\Pi, \varphi, M, D)$, then this can be shown by appeal to the foregoing rules. Likewise, if $\mathcal{F}(\Pi, \varphi, M, D)$, then this can be shown by appeal to the foregoing rules.

We shall use the usual notation $\bar{\bar{D}}$ for the cardinality of D . Note that when D is infinite, the rules $(\forall \mathcal{V})$ and $(\exists \mathcal{F})$ call for the construction of $\bar{\bar{D}}$ -furcating trees as, respectively, M -relative truth-makers for universals, and M -relative falsity-makers for existentials. So our metatheory needs to furnish such objects as needed.

14.11 SOME RESULTS; AND DISCUSSION

Metatheorem 1 *Modulo a metatheory which contains the mathematics of $\bar{\bar{D}}$ -furcating trees of finite depth, we have, for all models M with domain D ,*

$$\exists \Pi \mathcal{V}(\Pi, \varphi, M, D) \Leftrightarrow \varphi \text{ is true in } M$$

where the right-hand side is in the sense of Tarski.²⁷

²⁷ We abbreviated this right-hand side as $M \models \varphi$ in section 14.8.

Proof. The proof is by the obvious induction on the complexity of the sentence (or saturated formula) φ . Note that the proof is intuitionistic, provided only that the M -relative truth-makers for universals and the M -relative falsity-makers for existentials, in the case where D is infinite, can be assumed to exist, courtesy of the background metamathematics. This observation affects the right-to-left direction in the relevant cases of the inductive step.

Corollary 1 *Logical consequence \models in the Tarskian sense of Definition 1 coincides with logical consequence \models_{T} in the sense of Definition 3 in terms of quasi-effective transformability of truth-makers, but with “quasi-effective” interpreted as imposing no restriction at all on the Skolem function involved.*

Proof. Suppose $\varphi_1, \dots, \varphi_n \models \psi$. Let D be an arbitrary domain, and let M be a model with domain D . Let π_1, \dots, π_n be M -relative truth-makers for $\varphi_1, \dots, \varphi_n$ respectively. By Metatheorem 1, $\varphi_1, \dots, \varphi_n$ are true in M in the Tarskian sense. Hence ψ is true in M in the Tarskian sense. By Metatheorem 1 again, there is an M -relative truth-maker for ψ . Choose one such, by whatever method will ensure uniqueness of choice. Take that choice as the value $f(\pi_1, \dots, \pi_n)$. We thereby “construct” (in a possibly non-effective manner) a function f that can serve, in accordance with Definition 3, for the conclusion that $\varphi_1, \dots, \varphi_n \models_{\text{T}} \psi$.

For the converse, suppose $\varphi_1, \dots, \varphi_n \models_{\text{T}} \psi$. By Definition 3, let f be the function that will transform, for any model M , M -relative truth-makers for $\varphi_1, \dots, \varphi_n$ into an M -relative truth-maker for ψ . Let M' be any model in which $\varphi_1, \dots, \varphi_n$ are true in M' in the Tarskian sense. By Metatheorem 1, there are M' -relative truth-makers π_1, \dots, π_n , say, for $\varphi_1, \dots, \varphi_n$. Hence $f(\pi_1, \dots, \pi_n)$ is an M' -relative truth-maker for ψ . By Metatheorem 1 again, ψ is true in M in the Tarskian sense. Hence $\varphi_1, \dots, \varphi_n \models \psi$. QED

When the domain D is infinite, the furcations within truth-makers and falsity-makers that are called for in connection with the rules $(\exists\mathcal{F})$ and $(\forall\mathcal{V})$ will involve exactly as many branches as there are individuals in the domain. Such infinitary trees are not, of course, surveyable; but they are well-defined mathematical objects that can serve as representations of truth-makers and of falsity-makers for quantified sentences speaking of the infinitely many individuals in D . They should be congenial to the metaphysical and semantic realist, who entertains no skeptical qualms about the existence of infinitary objects, or about determinate truth-values for claims involving quantification over unsurveyable domains.

Thus far, no formal definition has been given in the truth-making literature, by philosophical logicians, for this quasi-technical notion of analytic metaphysics.²⁸ It is here

²⁸ Perhaps the best example of construction coming quite close to being truth-makers in our sense would be the infinitary proofs proposed by Carnap for arithmetic using the so-called ω -rule

$$\frac{\Psi 0 \quad \Psi s0 \quad \Psi ss0 \quad \dots \quad \Psi \underline{n}}{\forall n \Psi n}$$

proposed that the pre-formal notion of truth-maker (resp., falsity-maker) is nicely explicated by the foregoing formal notion of a truth-maker (resp., falsity-maker) of a saturated formula φ , *modulo* a domain D of discourse, from the basic facts represented in a complete and coherent set M of saturated literals.

The notion of a truth-maker for φ enables one also to isolate those possibly “smaller parts of reality” that might suffice to determine the truth of φ . Not all of M need be used, nor all of D surveyed, in order to determine the truth-value of φ . The exact “basic materials” on which that truth-value supervenes, so to speak, are captured by the truth-makers (or falsity-makers) available.

There could well be many different truth-makers for one and the same saturated formula under one and the same interpretation (i.e. *modulo* one and the same domain D , and one and the same set M of saturated literals). These different truth-makers represent different “constitutive routes” to the same truth-value under the interpretation in question. (Similarly for falsity-makers.)

14.11.1 Can we constrain the method of truth-maker transformation without changing the relation of classical logical consequence?

We were originally inspired by an interesting analogy with Prawitz’s definition of intuitionistic logical consequence, when suggesting that we define a more classical notion of consequence in terms of what we called “quasi-effective” transformability of truth-makers for the premises of an argument into a truth-maker for its conclusion. We have seen (Corollary 1) that we obtain exactly the relation of classical logical consequence if we impose no restriction at all on the kind of method used for the transformation of truth-makers. In effect, Corollary 1 established that

$$\begin{aligned} \varphi_1, \dots, \varphi_n &\models \psi \\ \Leftrightarrow \\ \exists f \forall M \forall \pi_1 \dots \forall \pi_n [&((\mathcal{V}_M(\pi_1, \varphi_1) \wedge \dots \wedge \mathcal{V}_M(\pi_n, \varphi_n)) \Rightarrow \mathcal{V}_M(f(\pi_1, \dots, \pi_n), \psi)) \end{aligned}$$

where the quantification $\exists f$ is over (set-theoretic) functions *tout court*.

which would be the \mathbb{N} -relative evaluation rule of “universal introduction” in our sense. It would need to be complemented, of course, by a dual rule for “existential elimination”:

$$\frac{\begin{array}{ccccccc} \frac{}{\Psi 0} \text{---}(i) & \frac{}{\Psi s 0} \text{---}(i) & \frac{}{\Psi ss 0} \text{---}(i) & & \frac{}{\Psi n} \text{---}(i) & & \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \\ \exists n \Psi n & \perp & \perp & & \perp & & \end{array}}{\perp} \text{---}(i)$$

The prima facie risk incurred by seeking to restrict the kind of function f that would be permissible here is that we might end up circumscribing too narrowly the resulting set of supposedly “classically valid” arguments $\varphi_1, \dots, \varphi_n : \psi$. The harder we make it to find such a function f , the more exigent a notion the definiendum \models becomes.

This exigency can be offset, however, by *more narrowly circumscribing* the class of models M with respect to which the f -transformations of truth-makers have to be effected. Noting that the quantification $\forall M$ is really a gloss for

$$(\forall \text{ domains } D)(\forall \text{ models } M \text{ based on domain } D)$$

we can inquire whether we might be able to constrain the kind of domain D involved, so as to offset any restriction of “quasi-effectiveness” that might be imposed on the function f . Obvious candidates are:

1. D is countable;
2. D is effectively enumerable; and/or
3. D is decidable.

Indeed, we might even consider

4. D is the set of natural numbers.

The challenge, as the present author sees it, is to find some formal delimitation of the class of permissible functions f that captures the orderliness of the transformations that are involved when one takes any *classical* natural deduction

$$\frac{\varphi_1, \dots, \varphi_n}{\Pi} \quad \psi$$

and accumulates on its premises respective M -relative truth-makers π_1, \dots, π_n :

$$\frac{\pi_1 \quad \dots \quad \pi_n}{\varphi_1 \quad \dots \quad \varphi_n} \quad \Pi \quad \psi$$

and thereupon seeks to “normalize” the resulting construction so that it becomes an M -relative truth-maker for ψ :

$$\frac{\pi_1 \quad \dots \quad \pi_n}{\varphi_1 \quad \dots \quad \varphi_n} \quad \Pi \quad \psi \quad \text{normalizes to} \quad \frac{\pi}{\psi}, \quad \text{where } \pi = v(\pi_1, \dots, \pi_n, \Pi)$$

We have seen displays like these before, of course, in section 14.8. But now it has to be borne in mind that the “lower” proof Π in this context is *classical*; and that the “upper” constructions π_1, \dots, π_n are (M -relative) *truth-makers*, not finitary M -warrants.

Here is a suggestion, followed by a conjecture. Seek some general but interestingly restrictive characterization Ξ of the transformation-functions v that are involved in this kind of normalization process—“general” enough to encapsulate them all, but “restrictive” enough to bring out the Skolemite flavor that was stressed above as providing illumination of the idea that the truth of the conclusion of a valid argument is somehow “contained in” the truth of its premises. The guess is that Ξ will characterize some class of functions reasonably low in the arithmetical hierarchy.

Now recall Prawitz’s soundness-via-normalization result, and its converse—his completeness conjecture—involving the relations \models_I and \models_P . We submit that there are analogues to be had for each of these.

To see this, recall that we have just talked above of a normalization process involving *classical* natural deductions, and truth-makers for their undischarged assumptions.

First, on the assumption that Ξ has been chosen so as to capture correctly the kind of method implicit in this normalization process, we shall have an analogue, for the classical case, of Prawitz’s soundness theorem. And this will be the case regardless of any restriction that might be imposed on the domains D —in fact, the stronger any such restriction might be, the easier it will be for functions f with property Ξ to effect what is required of them.

We propose imposing the restriction that the domain D be *decidable*. Of the four options listed above, this one strikes the present author as the most natural. So the normalization result that constitutes the soundness theorem for classical natural deductions will be stated as follows:

$$\begin{aligned} & \varphi_1, \dots, \varphi_n \vdash_C \psi \\ & \Rightarrow \\ & \exists f (\Xi f \wedge \forall \text{ decidable domains } D \forall \text{ models } M \text{ based on } D \\ & \forall \pi_1 \dots \forall \pi_n [((\mathcal{V}_M(\pi_1, \varphi_1) \wedge \dots \wedge \mathcal{V}_M(\pi_n, \varphi_n)) \Rightarrow \mathcal{V}_M(f(\pi_1, \dots, \pi_n), \psi))] \end{aligned}$$

That normalizability will hold, in the requisite sense, is made more plausible by Metatheorem 3 below.

Secondly, we venture to assert the converse—now in the form of a completeness conjecture. It is to be hoped that the proposed (and as yet unspecified) restriction Ξ on our transformation methods f will exactly offset the restriction that the domain D be decidable.²⁹

²⁹ We know already by the downward Löwenheim-Skolem Theorem that restricting domains D to be both countable and decidable will not result in any unwanted curtailment of the classical logical consequence relation.

Completeness conjecture for classical natural deduction:

$$\begin{aligned}
& \exists f(\exists f \wedge \forall \text{decidable domains } D \ \forall \text{models } M \text{ based on } D \\
& \forall \pi_1 \dots \forall \pi_n [((\mathcal{V}_M(\pi_1, \varphi_1) \wedge \dots \wedge \mathcal{V}_M(\pi_n, \varphi_n)) \Rightarrow \mathcal{V}_M(f(\pi_1, \dots, \pi_n), \psi))] \\
& \Rightarrow \\
& \varphi_1, \dots, \varphi_n \vdash_C \psi
\end{aligned}$$

We therefore commend to the reader the research problem of formulating a suitable property Ξ and establishing (or making highly plausible) the ensuing completeness conjecture in the form just given. One has to confront the problem, of course—which confronts Prawitz’s original completeness conjecture for the intuitionistic case as well—that the use of any *informal* terms such as “effective” and “decidable” render the conjecture incapable of truly *formal* proof. (In the same way, Church’s Thesis to the effect that every effective function on the naturals is recursive is incapable of truly formal proof.) But it is to be hoped that, despite this, some sort of “progress in persuasion” might be made, concerning both Prawitz’s completeness conjecture for the intuitionistic case, and the current completeness conjecture for the classical case that is modeled, by analogy, upon it.

14.11.2 Why falsity-makers?

Armstrong has raised the question (in personal correspondence):

A quick and unconsidered reaction is that a falsitymaker would be a truthmaker for the contradictory of the proposition made false. So why not use this truthmaker as a substitute for the fals[ity]maker?

This suggestion is a very reasonable first reaction. The challenge, however, is to characterize truth-makers for sentences φ *recursively*, unraveling the logical structure of φ . So there has to be a clause for negation. We need a clause along the lines of

$$\Pi \text{ is a truthmaker for } \neg\psi \text{ if and only if } \dots \psi \dots$$

where the blanks have to be filled in.

Compare this with

$$T \text{ is the truth value for } \neg\psi \text{ if and only if } F \text{ is the truth value for } \psi.$$

With two-valued truth-tables, we move easily between T and F , never entertaining the possible constraint (analogous to Armstrong’s suggestion above) that we should work only with the truth value T . In the truth-tabular context, Armstrong’s suggestion would be to the effect that we should regard any sentence ψ as having F for its truth value just in case T is the truth value for its negation $\neg\psi$. But it is the latter condition

(T is the truth value for $\neg\psi$) for which one seeks a “recursively unwinding” equivalent that will remove the negation sign; and that is what forces the recourse to the truth value F (for ψ).

Likewise with truth-makers (in place of assigning T) and falsity-makers (in place of assigning F).

Truth consists in the existence of a truth-maker; and falsity consists in the existence of a falsity-maker. The usual Frege-Tarski clause

$$\neg\psi \text{ is true if and only if } \psi \text{ is false}$$

becomes

$$\neg\psi \text{ has a truthmaker if and only if } \psi \text{ has a falsitymaker,}$$

or

$$\exists \Pi \Pi \text{ makestrue } \neg\psi \text{ if and only if } \exists \Sigma \Sigma \text{ makesfalse } \psi.$$

Now the temptation is irresistible to forge some kind of relation between satisfiers of the left-hand side of this biconditional, and satisfiers of its right-hand side. That is, we want to express the biconditional parametrically, without the existential quantifiers, in a form like

$$\Pi \text{ makestrue } \neg\psi \text{ if and only if } f(\Pi) \text{ makesfalse } \psi$$

—or, with the functional dependency going in the other direction,

$$\Pi \text{ makesfalse } \psi \text{ if and only if } g(\Pi) \text{ makestrue } \neg\psi.$$

Here, now, we can see that the function f would be an extraction, from within any truth-maker Π for $\neg\psi$, of a falsity-maker for ψ ; and that the function g would be an elaboration, of any falsity-maker Π for ψ , into a truth-maker for $\neg\psi$.

Both f and g are obviously furnished by the rule ($\neg\mathcal{V}$) above for constructing truth-makers for negations, to wit

$$\text{if } \Pi \text{ makesfalse } \psi, \text{ then } \frac{\Pi}{\neg\psi} \text{ makestrue } \neg\psi.$$

The converse of this conditional is in effect secured by the closure clause for the co-inductive definition of truth-makers and falsity-makers.

So now, if one is given a falsity-maker for ψ , one knows how to turn it into a truth-maker for $\neg\psi$: just apply a terminal step of the “negation introduction” rule ($\neg\mathcal{V}$). That is the operation g mentioned above. And, if one is given a truth-maker for $\neg\psi$, one knows how to obtain a falsity-maker for ψ : just extract its immediate subordinate construction (the one needed for the terminal application of ($\neg\mathcal{V}$)). That is the operation f mentioned above.

14.11.3 The question of the necessitation of truth by the existence of a truth-maker

We noted in section 14.4 that on our account, any truth-maker Θ for a sentence φ is so constituted that the resulting truthmaking is an *internal* relation, in the sense of Armstrong (2004). Here we clarify how this is so. As Alex Oliver has posed the problem,³⁰ we have to show how it is the case that

... the existence of one of [the] truthmakers [as here characterized] entails a truth it makes true. How could that be, given that [these] truthmakers are mathematical objects and can presumably exist even when the sentences they make true are false?

Our answer is as follows. Each truth-maker Θ , if it contains any “domain-fold” branchings (i.e. steps of $(\forall\mathcal{V})$ or $(\exists\mathcal{F})$) must, at each such step, have exactly one subordinate construction for each individual in the domain (D , say). So the existence of this kind of mathematical object Θ necessitates the truth of its conclusion φ when φ is interpreted with its quantifiers ranging over the domain D , and (for any model M with domain D that affords the atomic M -rules that have been used in the truth-maker) as making the relevant claim about M . This, we submit, is about as close as the truth-maker theorist of the current stripe can come to satisfying the Armstrongian demand that the existence of the truth-maker Θ should necessitate the truth of the sentence φ that it purports to make true. There is actually a nice air of generality about the implicit claim: Θ necessitates the truth of φ in *any* model with domain D that affords the atomic rules that have been used in Θ . It is as if the truth-maker theorist can say,

Necessarily, in any world like *this*: ...

[here pointing to the atomic M -rules used]

and with exactly *these individuals*: ...

[here indicating the range D of instantiating individuals for each and every step of $(\forall\mathcal{V})$ or $(\exists\mathcal{F})$ within Θ]

this truth-maker: ...

[here pointing to Θ]

makes-true the claim that φ .

³⁰ Personal correspondence.

14.11.4 The threatened slide to realism, and how to resist it

We foreshadowed earlier that our treatment of truth-makers might provide the realist with some new resources in his debate with the anti-realist. It is time now to reveal how this is so, and also to indicate how this new line of realist argument might be resisted.

The \mathcal{V} - and \mathcal{F} - rules are a straightforward transcription, into a recursive recipe for building constructions of the appropriate kinds, of the familiar inductive clauses in the Tarskian definition of truth or satisfaction. (This observation lies behind the obvious proof of Metatheorem 1.) Since the latter clauses, as already stressed, are acceptable to the anti-realist, it might be thought that the \mathcal{V} - and \mathcal{F} - rules would be acceptable also.

On the assumption (for the time being) that they are indeed acceptable, let us explore some further consequences.

Metatheorem 2 (Principle of Bivalence) *For all ϕ , either $\exists \Pi \mathcal{V}(\Pi, \phi, M, D)$ or $\exists \Sigma \mathcal{F}(\Sigma, \phi, M, D)$.*

Proof, and running anti-realist commentary. The proof is reasonably obvious (by induction on the complexity of ϕ), but it is important to realize that it is strictly classical.

The basis step of the inductive proof requires the completeness of M , in the sense explained above. This is the first point at which a classical conception obtrudes.

Classical metareasoning obtrudes further, at two more places.

First, in the inductive step dealing with saturated formulae of the form $\forall x \psi x$ we need, in the metalogic, the assurance that

either every D -instance $\psi \alpha$ has an M -relative truth-maker, or some D -instance $\psi \alpha$ does not have any M -relative truth-maker.

The first disjunct will allow one to construct an M -relative truth-maker for $\forall x \psi$ from the M -relative truth-makers for all its instances. By appeal to the Inductive Hypothesis the second disjunct will imply

some D -instance $\psi \alpha$ has an M -relative falsity-maker,

and this will allow one to construct an M -relative falsity-maker for $\forall x \psi$. Thus we obtain either an M -relative truth-maker or an M -relative falsity-maker for $\forall x \psi$.

Secondly, in the inductive step dealing with saturated formulae of the form $\exists x \psi x$ we need, in the metalogic, the assurance that

either every D -instance $\psi \alpha$ has an M -relative falsity-maker, or some D -instance $\psi \alpha$ does not have any M -relative falsity-maker.

The first disjunct will allow one to construct an M -relative falsity-maker for $\exists x\psi$ from the M -relative falsity-makers for all its instances. By appeal to the Inductive Hypothesis the second disjunct will imply

some D -instance $\psi\alpha$ has an M -relative truth-maker,

and this will allow one to construct an M -relative truth-maker for $\exists x\psi$. Thus we obtain either an M -relative falsity-maker or an M -relative truth-maker for $\exists x\psi$. *QED*

In handling these two quantifier clauses in our co-inductive definition of truth-makers and falsity-makers, the assurance that is needed has the logical form

$$\forall x\Psi x \vee \exists x\neg\Psi x.$$

If this is taken to be available as a theorem of the metalogic, then we are dealing with a *strictly classical* metalogic.³¹ For this principle is *not* acceptable to the intuitionist. So the intuitionist is unable (at least, by *this* method of proof) to deliver Metatheorem 2—which is as it should be.

Metatheorem 3 *Any M -relative truth-maker for $\neg\neg\phi$ contains an M -relative truth-maker for ϕ . Hence in the object language we have $\neg\neg\phi \models \phi$.*

Proof. Suppose Π is an M -relative truth-maker for $\neg\neg\phi$. According to the closure clause of our co-inductive definition of truth-makers and falsity-makers, Π must have been constructed by appeal to the rule $(\neg\mathcal{V})$, and accordingly have the form

$$\frac{\frac{\text{---}}{\neg\phi}^{(i)} \quad \Sigma}{\perp} \quad \frac{}{\neg\neg\phi}^{(i)}$$

where Σ is an M -relative falsity-maker for $\neg\phi$. According to the closure clause of our co-inductive definition of truth-makers and falsity-makers, Σ must have been constructed by appeal to the rule $(\neg\mathcal{F})$, and accordingly have the form

$$\frac{\Theta \quad \frac{\neg\phi \quad \phi}{\perp}}{\perp}$$

where Θ is an M -relative truth-maker for ϕ . *QED*

³¹ This point applies equally well to the case where the classically needed but intuitionistically disputed principle has the logical form $\forall x\neg\Psi x \vee \exists x\Psi x$.

The anti-realist's immediate objection to this result will involve challenging the framing of the co-inductive definition of truth-makers and falsity-makers. For the anti-realist envisages situations where it is known that $\neg\phi$ itself leads to absurdity, without that in itself guaranteeing that there is warrant for the assertion of ϕ . So the anti-realist's real complaint will be against the closure clause of the inductive definition. For this closure clause, particularly as it applies to falsity-makers for negations, makes the construction rule ($\neg\mathcal{F}$), with its built-in guarantee of a truth-maker for ϕ , the sole route to the rejection of a negation. While not having space here to develop this criticism on behalf of the anti-realist, we nevertheless wish at least to *identify* the exact issue that has to be joined in any further debate with the realist over this all-too-easy-looking validation of the strictly classical rule of double-negation elimination.

14.11.5 An important application, not hitherto stressed in the truth-making literature

The notion of a truth-maker (or falsity-maker) allows one to analyze what basic facts might be *relevant* to the truth (or falsity) of a saturated formula. To be sure, truth-makers and falsity-makers can be infinitary, but this is the case only when the domain D itself is infinite; and nothing less could be expected of a truth-maker for a universal claim (or of a falsity-maker for an existential one) in such situations.

Because truth-makers and falsity-makers allow us to focus more clearly on the “local” grounds for truth or falsity of a given sentence, they can be employed as important tools when explicating the notion of cognitive significance. A crucial explicandum in this project is the idea of *creative extension* of a given theory by the incorporation of new theoretical vocabulary involved in new hypotheses. In spelling out this idea, the need arises to keep track of how it will always be the case that facts expressible in the observational vocabulary will feature in the “grounds for falsity” of any hypothesis that is able to be refuted *by the observational evidence*. That is how the hypotheses in question acquire cognitive significance, and impart it to the new theoretical terms that they contain. It turns out that this idea is key—and sufficient—for avoiding the well-known Church-style collapses of the resulting formally defined notion of cognitive significance.³² For the purposes of explicating this idea of creative extension and the necessary involvement of observational facts, “whole models” M prove to be too blunt or crude. It turns out that what is needed, instead, are the foregoing notions of M -relative truth-makers and falsity-makers. For a fuller development of these ideas, and an “adequacy metatheorem” concerning the resulting notion of cognitive significance, the reader is referred to Tennant (1997: ch. 11).

³² See Church's famous review (Church 1949) of Ayer (1946), in which Church showed how to trivialize Ayer's notion of indirect verifiability (which Ayer had offered as his formal explication of what we call cognitive significance).

14.11.6 The question of language-independence

A truth-maker theorist who wishes to define truth-makers in such a way as to ensure that they are *language-independent* might take issue with our appeal to predicate-symbols that belong to a particular language.³³ But it is of course open to such a theorist to secure the complete language-independence of truth-makers by stipulating (as for example Prawitz 1968) that any n -place predicate-symbol R is to be interpreted, at all its occurrences within labels of nodes of a truth-maker, as the (language-independent) n -place *relation* R itself (in the unary case, as REDNESS, say, rather than as "... is red" or "... est rouge").

Note that the inductive clauses in the definition of truth-makers and falsity-makers, as labeled trees, can be interpreted in such a way as to make the left-right ordering of immediate subtrees immaterial to the identity of the tree overall. That is to say, the immediate subtrees need not be taken as given in any particular order. This is especially the case with the possibly infinitary quantifier rules ($\exists\mathcal{F}$) and ($\forall\mathcal{V}$), in which the premises for the applications of those rules can be construed as functions mapping individuals from the domain D to falsity-makers, or truth-makers, respectively, for the relevant instances of the quantified matrix.

REFERENCES

- Armstrong, D. M. (2004). *Truth and Truthmakers*. Cambridge: Cambridge University Press.
- Ayer, A. J. (1946). *Language, Truth and Logic*, 2nd edn, revised and re-set. London: Victor Gollancz.
- Beebe, H. and J. Dodd (eds.) (2005). *Truthmakers: The Contemporary Debate*. Oxford: Oxford University Press.
- Bigelow, J. (1988). *The Reality of Numbers: A Physicalist's Philosophy of Numbers*. Oxford: Clarendon Press.
- Church, A. (1949). "Review: Alfred Jules Ayer, *Language, Truth and Logic* [review of 2nd edn]. *Journal of Symbolic Logic* 14: 52–3.
- Dummett, M. (1976). "What is a Theory of Meaning? II," in G. Evans and J. McDowell (eds.), *Truth and Meaning*. Oxford: Oxford University Press, 67–137.
- Fox, J. F. (1987). "Truthmaker," *Australasian Journal of Philosophy* 65: 188–207.
- Frege, G. (1884; reprinted 1961). *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl*. Hildesheim: Georg Olms.
- Gentzen, G. (1934–35). "Untersuchungen über das logische Schliessen," *Mathematische Zeitschrift* 1, 2: 176–210, 405–431. Translated as "Investigations into Logical Deduction," in *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo (1969). Amsterdam: North-Holland, 68–131.
- Lowe, E. J. and A. Rami (eds.) (2009). *Truth and Truth-Making*. Ithaca: McGill-Queen's University Press.

³³ The author is grateful to Kit Fine for raising this problem.

- Mulligan, K., P. Simons, and B. Smith (1984). "Truth-Makers," *Philosophy and Phenomenological Research* 44: 287–321.
- Prawitz, D. (1968). "Propositions," *Theoria* 34: 134–46.
- Prawitz, D. (1974). "On the Idea of a General Proof Theory," *Synthese* 27: 63–77.
- Restall, G. "Truth-makers, Entailment and Necessity," *Australasian Journal of Philosophy* 74: 331–40.
- Sundholm, G. (1994). "Existence, Proof and Truth-Making: A Perspective on the Intuitionistic Conception of Truth," *Topoi* 13: 117–26.
- Tarski, A. (1956) [1936]. "On the Concept of Logical Consequence," in J. H. Woodger (ed.), *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press, 409–20.
- Tarski, A. (1956) [in Polish, 1933]. "The Concept of Truth in Formalized Languages," in J. H. Woodger (ed.), *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press, 152–278.
- Tennant, N. (1978). *Natural Logic*. Edinburgh: Edinburgh University Press.
- Tennant, N. (1979). "Entailment and Proofs," *Proceedings of the Aristotelian Society* 79: 167–89.
- Tennant, N. (1987). *Anti-Realism and Logic: Truth as Eternal*. Oxford: Clarendon Library of Logic and Philosophy, Oxford University Press.
- Tennant, N. (1992). *Autologic*. Edinburgh: Edinburgh University Press.
- Tennant, N. (1997). *The Taming of The True*. Oxford: Oxford University Press.
- Tennant, N. (2008). "Natural Logicism via the Logic of Orderly Pairing," in S. Lindström, E. Palmgren, K. Segerberg, and V. Stoltenberg-Hansen (eds.), *Logicism, Intuitionism, Formalism: What has Become of Them?* Synthese Library 341. Dordrecht: Springer, 91–125.
- Tennant, N. (2010). "Inferential Semantics," in J. Lear and A. Oliver (eds.), *The Force of Argument: Essays in Honor of Timothy Smiley*. London: Routledge, 223–57.
- Tennant, N. (2012). *Changes of Mind: An Essay on Rational Belief Revision*. Oxford: Oxford University Press.
- Tennant, N. (2015). "A New Unified Account of Truth and Paradox," *Mind* 124: 571–605.

CHAPTER 15

BIVALENCE AND DETERMINACY

IAN RUMFITT

THE Principle of Bivalence is the thesis that every statement is either true or false. By a statement, I mean a declarative utterance or inscription that expresses a complete thought. Statements, in this sense, need not be assertions: the antecedent of the conditional “If Tom is in Germany, he is in Berlin” expresses the thought that Tom is in Germany and hence qualifies as a statement, even though a speaker who affirms the whole conditional does not assert that Tom is in Germany. The notion of expressing a complete thought is not completely clear and, as what follows will show, the truth of the Principle depends, among other factors, on how it is clarified. But it is clear enough to identify our topic.

The Principle of Bivalence has been a bone of philosophical contention for centuries. An apparently powerful argument for it sits alongside apparently convincing counterexamples to it. I start with the classic argument for the Principle.

15.1 THE SIMPLE ARGUMENT FOR BIVALENCE

The argument I have in mind is all but explicit in Aristotle. Everybody remembers Aristotle’s explanations of truth and falsity in *Metaphysics* Γ.7, but it is sometimes forgotten that they form part of an argument for Bivalence:

Of one subject we must either affirm or deny any one predicate. This is clear, in the first place, if we define what the true and the false are. To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true; so that he who says of anything that it is, or that it is not, will say either what is true or what is false (*Metaphysics* 1011b25; cf. Plato, *Cratylus* 385b2).

The idea behind these definitions is this: an utterance is true if it says that things are somehow, and they are thus; it is false if it says that things are somehow, and they are not thus. The non-nominal quantifier “somehow” is one that English speakers understand, but in tracing out the implications of the proposed definitions, we shall need a rigorous statement of the rules governing its use. What are those rules?

In his book *Objects of Thought* (1971), Arthur Prior gave an answer. Let us suppose that a standard first-order formalized language is enriched with universal and existential quantifiers whose attendant variables replace complete well-formed formulae. Let us also suppose that the introduction and elimination rules for these two new “propositional” quantifiers are analogues of the rules that govern the corresponding quantifiers into name position, aside from differences consequential upon the different syntactic categories of the associated variables. Then, for example, assuming that the underlying propositional logic is classical, the formulae “ $\forall P (P \vee \neg P)$ ” and “ $\neg \exists P (P \wedge \neg P)$ ” will be logical theorems. What Prior noticed is that if we read “ $\forall P$ ” as “However things may be said or thought to be,” read “ $\exists P$ ” as “There is a way things may be said or thought to be,” and read the associated variables as “they are thus,” then the theorems of this system emerge as logical truths under the proposed interpretation. Thus the theorem “ $\forall P (P \vee \neg P)$ ” says, “However things may be said or thought to be, either they are thus or they are not thus,” which a classical logician will take to be a logical truth. Reversing the translation, we can formalize Aristotle’s definitions of truth and falsity as follows, where “ $\forall u$ ” is an ordinary objectual quantifier ranging over utterances (which I shall henceforth take to include inscriptions), and “ $\exists P$ ” is the existential propositional quantifier lately explained:

(T) $\forall u(\text{True}(u) \leftrightarrow \exists P(\text{Say}(u, P) \wedge P))$

(F) $\forall u(\text{False}(u) \leftrightarrow \exists P(\text{Say}(u, P) \wedge \neg P))$.

(T) and (F), then, capture Aristotle’s definitions of truth and falsity in a way that enables us rigorously to trace out their implications.¹

Aristotle presents his formulae as definitions, and (T) and (F), if they are correct at all, can serve as explicit definitions of “true” and “false” as these notions apply to utterances.² Just for this reason, though, it might seem as if (T) and (F) cannot be correct: they appear to conflict with Alfred Tarski’s theorem that truth is undefinable (Tarski 1935). Tarski’s theorem applies to a formalized language, L , containing a

¹ Read strictly, Aristotle affirms only the right-to-left halves of (T) and (F), but the converse conditionals are implied by his advertising his formulae as *definitions* of truth and falsity. If I say “This result is clear if we define what an Abelian group is. If a group is commutative it is Abelian,” my second sentence has the force of a biconditional.

² Corresponding formulae could serve as definitions of truth and falsity as these notions apply to states of belief and to other bearers of truth and falsehood. For some discussion of the range of truth-bearers, see Rumfitt (2011: section 2).

negation operator \neg and a device $\ulcorner \cdot \urcorner$ which, when applied to any well-formed expression of L , yields a singular term designating that expression. It is assumed that the syntax of the language is strong enough to prove the diagonal lemma: for any formula $A(x)$ in L , with x free, there is a formula B in L such that B is equivalent to $A(\ulcorner B \urcorner)$. According to Tarski, a *truth-predicate* for L is a one-place predicate $Tr(\xi)$ such that, for any closed formula A of L , $Tr(\ulcorner A \urcorner)$ is equivalent to A : Tarski assumes that L contains no context-sensitive expressions, which is why he can take a truth-predicate for L to be a unary predicate of L 's type sentences. What Tarski then proves is that no truth-predicate for L can be a predicate in L . For suppose $Tr(\xi)$ were a truth-predicate for L in L . Then $\neg Tr(x)$ would be a well-formed formula of L with x the only free variable. By the diagonal lemma, there would exist a closed formula D in L such that D is equivalent to $\neg Tr(\ulcorner D \urcorner)$. Since $Tr(\xi)$ is a truth-predicate for L , we would also have that $Tr(\ulcorner D \urcorner)$ is equivalent to D , so that $Tr(\ulcorner D \urcorner)$ and $\neg Tr(\ulcorner D \urcorner)$ would be equivalent. This contradiction reduces to absurdity the supposition that $Tr(\xi)$ is a truth-predicate in L . So no truth-predicate for L can be in L (Tarski 1935: 249–51). Although Aristotle does not of course attain modern standards of explicitness in such matters, he does seem to advance, in Greek, an account of a truth-predicate which applies (among other things) to Greek utterances and inscriptions, so Tarski's theorem appears to cast doubt on his account. It also casts doubt on (T) and (F), insofar as the predicates they purport to define are predicates of a semi-formalized English which apply, inter alia, to utterances and inscription in that semi-formal language.

Given Tarski's assumptions, his proof of the indefinability theorem is unassailable, but one of those assumptions is highly contestable. Tarski assumes that if $Tr(\xi)$ is a truth-predicate for L , then any closed formula A of L is equivalent to $Tr(\ulcorner A \urcorner)$. He thereby presupposes that any closed formula of L has truth-conditions. A formula will have truth-conditions, though, only if it expresses a complete thought, so Tarski's argument presupposes that any closed formula of the relevant language expresses such a thought. The very formula that Tarski uses in proving his theorem, however, casts doubt on that presupposition. That formula "says of itself" that it is not true. To see why Tarski's presupposition is doubtful, then, let us consider a corresponding English inscription, λ :

The inscription seven lines from the foot of p. 396 of this book is not true.

When applied to λ , (T) yields

- (1) $\text{True}(\lambda) \leftrightarrow \exists P(\text{Say}(\lambda, P) \wedge P)$.

Now if λ says anything, what it says is that the inscription seven lines from the foot of p. 396 of this book is not true, so on the assumption that λ does succeed in saying something, (1) yields

- (2) $\text{True}(\lambda) \leftrightarrow$ the inscription seven lines from the foot of p. 396 of this book is not true.

By inspection we also have

- (3) λ = the inscription seven lines from the foot of p. 396 of this book.

By Leibniz's Law, (2) and (3) together entail

- (4) $\text{True}(\lambda) \leftrightarrow \neg \text{True}(\lambda)$.

As in Tarski's proof of his indefinability theorem, (4) is a contradiction, but we need not take this as reducing to absurdity the assumption that semi-formal English contains a predicate with the intended sense of "true." In order to move from (1) to (2), we need the assumption that λ expresses a thought and we may, instead, take the contradiction to refute that assumption. Tarski entirely overlooks this possibility. He writes that we "wish to use the term 'true' in such a way that *all* [T-equivalences in the form 'S is true if and only if P'] can be asserted, and we shall call a definition of truth 'adequate' if [and only if] *all* these equivalences follow from it" (Tarski 1944: 344; emphasis added). But what this sort of case reveals is that our commitment to assert such T-equivalences is provisional. We shall assert "S is true if and only if P" only when we believe that S expresses a thought; in particular, we shall assert (2) only when we believe this of λ . For all that Tarski says, the above deduction is simply a proof by *reductio* that λ fails to express a thought, i.e. that $\neg \exists P \text{ Say}(\lambda, P)$. On this view, (T) and (F) transform "the semantic paradoxes . . . into sound arguments for constraints on what can say what in what contexts" (Williamson 1998: 19).

Whether this approach gives the best solution to the paradox the liar is a large issue that I cannot address here.³ It is clear, though, that the theory of truth and falsity that comprises (T), (F), and their logical consequences is formally consistent. As Timothy Williamson has observed, we can show this "by constructing an unintended model . . . in which formulas are treated as referring to truth-values, the propositional quantifiers range over truth-values, and all formulas of the forms 'say (A, c, P)', 'True (A, c)', and 'False (A, c)' are treated as false" (Williamson 1998: 14).

Assuming that the logic of this theory is classical, the Principle of Bivalence seems, at first blush, to be a theorem of it. For, given an arbitrary statement, u , we can reason as follows:

- | | | |
|-----|-------------------------------|----------------------------------|
| (1) | $\exists P \text{ Say}(u, P)$ | Definition of statement |
| (2) | $\text{Say}(u, P)$ | (1), existential instantiation |
| (3) | $P \vee \neg P$ | Excluded middle |
| (4) | P | Assumption |
| (5) | $\text{Say}(u, P) \wedge P$ | (2), (4), \wedge -introduction |

³ For approaches to the Liar along these general lines, see Prior (1971: ch. 6); Kneale (1972); Mackie (1973: ch. 6); Parsons (1974); Smiley (1993); Glanzberg (2001); and Rumfitt (2014).

(6) $\exists P(\text{Say}(u, P) \wedge P)$	(5), existential generalization
(7) $\text{True}(u)$	(6), (T)
(8) $\text{True}(u) \vee \text{False}(u)$	(7), \vee -introduction
(9) $\neg P$	Assumption
(10) $\text{Say}(u, P) \wedge \neg P$	(2), (9), \wedge -introduction
(11) $\exists P(\text{Say}(u, P) \wedge \neg P)$	(10), existential generalization
(12) $\text{False}(u)$	(11), (F)
(13) $\text{True}(u) \vee \text{False}(u)$	(12), \vee -introduction
(14) $\text{True}(u) \vee \text{False}(u)$	(3), (8), (13) \vee -elimination, with discharge of assumptions (4) and (9)

Since u was an arbitrarily chosen statement, we may generalize to reach the conclusion “Every statement is either true or false.” I shall call this deduction the *Simple Argument* for Bivalence. It spells out the argument that is implicit in the passage I quoted from Aristotle.

15.2 PROBLEMS WITH THE SIMPLE ARGUMENT: DETERMINACY OF SENSE

The Simple Argument may seem to be conclusive, at least if the underlying logic is classical. In fact, though, there is a serious problem with it.

The Principle of Bivalence, I said, is the thesis that every statement is either true or false, but the “or” here is exclusive, not the logician’s *vel*. That is, the Principle might more explicitly be formulated as “Every statement is either true or false but not both.” However, even in classical logic, which has “ $\neg\exists P(P \wedge \neg P)$ ” as a theorem, (T) and (F) leave open the possibility that a statement is both true and false. A statement might be true by virtue of saying that P in a circumstance where P , and at the same time false by virtue of saying that Q in a circumstance where not Q . Where “ P ” and “ Q ” are distinct formulae, the truth of “ $P \wedge \neg Q$ ” may well be a logical possibility. Even given classical logic, then, nothing in the Aristotelian definitions (T) and (F) excludes the possibility that a statement is both true and false.

What would exclude that possibility is the further assumption that the statement has a *determinate sense*. Any statement, we said, expresses a complete thought. But if it has a determinate sense, it will express just one such thought: there will be such a thing as *the* thought it expresses. While there may be alternative ways of expressing that thought, those ways cannot diverge in truth value. If u has a determinate sense, then, it will meet the following condition:

(D) $\forall P \forall Q (\text{Say}(u, P) \wedge \text{Say}(u, Q) \rightarrow (P \leftrightarrow Q)).$

If u has a determinate sense it will meet stronger conditions too, such as the following necessitated form of (D):

$$\Box(\forall P\forall Q(\text{Say}(u, P) \wedge \text{Say}(u, Q) \rightarrow \Box(P \leftrightarrow Q))).$$

In the present essay, though, I am not concerned with the modal aspects of statements, so we may focus on (D), which I shall take to be the defining condition for u to be *determinate*. Given (T) and (F), and the premise that u satisfies (D), it is straightforward to show that u cannot be both true and false.

How might a statement fail to be determinate? How could an utterance say something without there being one thing that it says? We should set aside as irrelevant a shallow sense in which this is possible. If I say “It is wet and cold,” I may be reported as having said that it is wet, and as having said that it is cold. In the intended sense, though, “Say(u, P)” means “The *whole* of what u says is that P ”; partial reports of an utterance’s content are beside the point.

Even in the intended sense of “Say(u, P),” however, (D) may fail. As we shall see in more detail below, special problems attend the use of classical logic when this is applied to statements involving unrestricted quantification over sets. The usual semantic justification for the use of classical logic when reasoning with first-order statements presumes that the domain of quantification constitutes a set. In standard set theory, however, there is no set of all sets. An alternative justification for the use of classical logic in set theory rests on the view that apparently unrestricted quantification over sets is implicitly restricted to the members of a standard model for set theory. Since there are many such models, however, this view casts doubt upon (D). For what, on this view, does the statement “There exists a strongly inaccessible cardinal” say? It “says” that there is such a cardinal in the smallest standard model V_{κ_1} ; but it equally “says” that there is such a cardinal in the second smallest model V_{κ_2} , and so forth. The claims that our statement “says” are not even materially equivalent. The first inaccessible, κ_1 , is a member of V_{κ_2} but not of V_{κ_1} so, on the first reading, the statement is false whereas it is true on the second reading. I put the word “say” in scare quotes because it is somewhat strained to use the term in this way: we want to ask, “Well, *which* does it say?” On the view in question, though, the only available sense for “ u says that P ” is that of “An interpretation of u under which it says that P is as legitimate as any other interpretation.” The existence of equally legitimate ways of interpreting u should not prevent us from classifying u as a statement: it is not that it fails to say anything. All the same, in such a case no *unique* content can be ascribed to it.

For statements that exhibit this kind of indeterminacy, (T) is inadequate as an account of truth. If the only available sense for “ u says that P ” is “An interpretation of u under which it says that P is as legitimate as any other interpretation” then the only available sense of the formula “ $\exists P(\text{Say}(u, P) \wedge P)$ ” is “An interpretation which renders u true is as legitimate as any other interpretation”; but to say so much is to say something weaker than “ u is true.” Similarly, (F) is inadequate as an account of falsity. When the

only available sense for “ u says that P ” is as above, “ $\exists P(\text{Say}(u, P) \wedge \neg P)$ ” means “An interpretation which renders u false is as legitimate as any other interpretation.” In such a case, then, the Simple Argument fails to establish that u is either true or false (even in the weak sense in which “or” means *vel*). Its conclusion says only that u is either such that it could, with maximal legitimacy, be taken to be true, or could, with equal legitimacy, be taken to be false.

How should we define truth and falsity when there is a serious possibility of some statements being indeterminate? The natural answer is this. We count a statement as true if, *however* it may legitimately be taken to say that things are, they are thus; we count it as false if, *however* it may legitimately be taken to say that things are, they are not thus. In other words, when indeterminacy of sense is a serious possibility we need to replace (T) and (F) by the following:

$$(T^*) \quad \forall u(\text{True}(u) \leftrightarrow (\exists P \text{ Say}(u, P) \wedge \forall Q(\text{Say}(u, Q) \rightarrow Q)))$$

$$(F^*) \quad \forall u(\text{False}(u) \leftrightarrow (\exists P \text{ Say}(u, P) \wedge \forall Q(\text{Say}(u, Q) \rightarrow \neg Q))).$$

Like (T) and (F), (T*) and (F*) may be regarded as explicit definitions.

The theory whose axioms are (T*) and (F*) in a logical system permitting quantification into sentence position is again consistent. In fact, this new theory is consistent even if every utterance in the domain of quantification says something. This may be shown by constructing an unintended interpretation in which each utterance both “says” that P and “says” that not P , so that every formula “ $\exists P \text{ Say}(u, P)$ ” is true, while every formula of the form “True(u)” or “False(u)” is false. In this new theory, the semantic paradoxes are transformed into sound arguments to show that a paradoxical utterance expresses no *unique* thought.

When we took truth and falsity to be defined by (T) and (F), we needed the assumption that u is determinate in order to show that u is not both true and false. Now that truth and falsity are defined by (T*) and (F*), we do not need the assumption of determinacy to show that truth and falsity are contraries. Under the new definitions, however, we need determinacy in order to show that u is true *vel* false. For now that “true” is defined by (T*), we cannot move from line (6) of the Simple Argument, *viz.* “ $\exists P(\text{Say}(u, P) \wedge P)$,” to line (7), *viz.* “True(u).” Similarly, now that “false” is defined by (F*), we are blocked from moving from line (11) to line (12). Given the assumption that u satisfies (D), however, we can recast the deduction as follows:

- | | | |
|-----|---|--------------------------------|
| (1) | $\exists P \text{ Say}(u, P)$ | Definition of statement |
| (2) | $\text{Say}(u, P)$ | (1), existential instantiation |
| (3) | $\forall Q(\text{Say}(u, P) \wedge \text{Say}(u, Q) \rightarrow (P \leftrightarrow Q))$ | (D), universal instantiation |
| (4) | $\forall Q(\text{Say}(u, Q) \rightarrow (P \leftrightarrow Q))$ | (2), (3) |
| (5) | $P \vee \neg P$ | Excluded middle |

(6)	P	Assumption
(7)	$\forall Q(\text{Say}(u, Q) \rightarrow Q)$	(4), (6)
(8)	$\exists P \text{Say}(u, P) \wedge \forall Q(\text{Say}(u, Q) \rightarrow Q)$	(1), (7) \wedge -introduction
(9)	$\text{True}(u)$	(8), (T^*)
(10)	$\text{True}(u) \vee \text{False}(u)$	(9), \vee -introduction
(11)	$\neg P$	Assumption
(12)	$\forall Q(\text{Say}(u, Q) \rightarrow \neg Q)$	(4), (11)
(13)	$\exists P \text{Say}(u, P) \wedge \forall Q(\text{Say}(u, Q) \rightarrow \neg Q)$	(1), (12) \wedge -introduction
(14)	$\text{False}(u)$	(13), (F^*)
(15)	$\text{True}(u) \vee \text{False}(u)$	(9), \vee -introduction
(16)	$\text{True}(u) \vee \text{False}(u)$	(5), (10), (15) \vee -elimination, with the discharge of assumptions (6) and (11)

(I have elided some elementary logical steps at lines (4), (7), and (12).) I shall call this latest deduction the *Revised Argument*. Its conclusion is a restricted version of Bivalence: every *determinate* statement is either true or false.

Miroslava Andjelković and Timothy Williamson (2000) claim to be able to deduce the thesis that every statement satisfies (D) from (T^*). If their argument worked, (T^*) would entail (T) and (F^*) would entail (F); moreover, the Simple Argument would apply to every statement and we would have no need for the Revised Argument. Andjelković and Williamson reason as follows. Where s and t are type sentences in a given language L , let sEt be the sentence formed by writing s , then a sign E in L that means “if and only if,” and finally t . (It is assumed that L contains at least one expression meaning “if and only if.”) sEt , they claim, will always express a biconditional proposition whose first component is what s says and whose last component is what t says. That is, where $\langle s, c \rangle$ is the (possible) utterance of sentence s in the context c ,

$$(E1) \quad \forall s \forall t \forall c \forall P \forall Q [\text{Say}(\langle s, c \rangle, P) \wedge \text{Say}(\langle t, c \rangle, Q) \rightarrow \text{Say}(\langle sEt, c \rangle, P \leftrightarrow Q)].$$

Andjelković and Williamson further claim that if the utterance of a sentence s in a context says something, then an utterance in that same context of a complex sentence in which an occurrence of s is followed by E , which in turn is followed by a second occurrence of s , is true:

$$(E2) \quad \forall s \forall c \forall P [\text{Say}(\langle s, c \rangle, P) \rightarrow \text{True}(\langle sEs, c \rangle)].$$

Now a special case of (E1) is

$$(1) \quad \forall s \forall c \forall P \forall Q [\text{Say}(\langle s, c \rangle, P) \wedge \text{Say}(\langle s, c \rangle, Q) \rightarrow \text{Say}(\langle sEs, c \rangle, P \leftrightarrow Q)].$$

Furthermore, (T^*) yields

$$(2) \quad \forall s \forall c \forall P \forall Q [\text{Say}(\langle sEs, c \rangle, P \leftrightarrow Q) \rightarrow (\text{True}(\langle sEs, c \rangle) \rightarrow (P \leftrightarrow Q))].$$

From (1) and (2) we get

$$(3) \quad \forall s \forall c \forall P \forall Q [\text{Say}(\langle s, c \rangle, P) \wedge \text{Say}(\langle s, c \rangle, Q) \rightarrow (\text{True}(\langle sEs, c \rangle) \rightarrow (P \leftrightarrow Q))].$$

But (E2) and (3) yield

$$(4) \quad \forall s \forall c \forall P \forall Q [\text{Say}(\langle s, c \rangle, P) \wedge \text{Say}(\langle s, c \rangle, Q) \rightarrow (P \leftrightarrow Q)].$$

Since any possible utterance is the utterance of a sentence in a context, (4) amounts to the claim that any possible statement satisfies (D).

How does this argument fare when the only available interpretation of “ u says that P ” is “An interpretation of u under which it says that P is as legitimate as any other interpretation”? When “says” has that meaning, (E2) and (1) cannot both be correct. (1) implies that, in a given context, a speaker-hearer may give non-equivalent, but equally legitimate, interpretations to the two occurrences of s in sEs . (E2), by contrast, presumes that those two occurrences will be interpreted in equivalent ways: if $\langle sEs, c \rangle$ is always true, then an utterance of sEs must either be heard as saying that $P \leftrightarrow P$ or be heard as saying that $Q \leftrightarrow Q$ (where P and Q are the legitimate interpretations of s). The latter presumption may well be correct: there may be a convention in the relevant language that precludes switching from one legitimate interpretation of s to another when interpreting the two occurrences of s in the complex sentence sEs . But if such a convention is in place, (1) will be false.

Andjelković and Williamson are right to say that (E1) logically entails (1). That point, however, merely shows that one’s first thought—that (E1) is a wholly exceptionless truth about biconditionals—may need to be revised. It is an interesting question how it is best emended. One suggestion is this. Let us call a context *regular* when it is understood that repeated occurrences of a single expression type in the same sentence must be interpreted in the same way.⁴ To ensure that (E2) is true, we must restrict the range of the variable “ c ” to regular contexts. When the range of “ c ” is so restricted, though, (E1) needs to be revised by adding the conjunct $s \neq t$ to its antecedent; this ensures that (1) is not among its consequences. In any case, whether or not this way of preserving what is right in (E1) is optimal, the Andjelković-Williamson argument fails. Indeterminate

⁴ I write of expressions in general, for puzzle cases of this kind are not confined to complete sentences. In any regular context, “Princeton = Princeton” will be true, even though “Princeton” could equally legitimately be taken to refer to Princeton Borough or to Princeton Township, and Princeton Borough \neq Princeton Township (David Lewis’s example; see Lewis 1988: 128).

statements are a real possibility and (*T*), (*F*), and the Simple Argument all need to be revised to take account of them.

The Revised Argument shows, indeed, from where challenges to Bivalence must come. A non-bivalent statement must either be indeterminate—i.e. fail to satisfy (*D*)—or express thoughts to which classical logic does not apply. In the latter case, the inapplicable laws will almost certainly be excluded middle (invoked at line (5)) or the rule of \vee -elimination (alias “proof by cases”) that is needed at the last step. One reason for thinking that this latest reconstruction of the argument for Bivalence is on the right lines is that the most interesting challenges to the bivalence of a given statement proceed by questioning its determinacy, or the application to it of excluded middle, or of proof by cases. With that in mind, I turn to consider three putative counterexamples to Bivalence in the light of our analysis.

15.3 FUTURE CONTINGENTS

The first of these putative counterexamples is not, I think, especially persuasive, but it needs to be mentioned because it looms so large in the history of the topic.

On the traditional reading, Aristotle put forward a counterexample to Bivalence in the ninth chapter of his early treatise *De interpretatione*.⁵ “A sea-battle will take place tomorrow” expresses a complete thought, and so qualifies as a statement. But to ascribe truth to that statement, Aristotle seems to argue, would imply that it is settled that battle will be joined; and to ascribe falsehood to it would imply that it is settled that battle will be avoided. So, at a time when it is not settled whether battle will be joined, we cannot say that the statement is either true or false.

Aristotle’s discussion of the case is convoluted and hard to interpret. Part of the difficulty stems from his switching between two formulations of Bivalence. We can see this switch in the passage already quoted from the *Metaphysics*. The passage concludes with a good formulation of the Principle:

- (1) He who says . . . anything . . . will say either what is true or what is false.

But it starts with a potentially misleading formulation:

- (2) Of one subject we must either affirm or deny any one predicate.

⁵ The traditional reading has been challenged. The debate continues, but classic contributions include Łukasiewicz (1922); Anscombe (1956); Kneale and Kneale (1962: ch. 2, section 4); Hintikka (1964); and von Wright (1984).

(2) may be read as equivalent to (1). However, mentioning the speech acts of affirmation and denial muddies the waters, for (2) is more naturally read as expressing a principle that is not equivalent to Bivalence:

(3) Given any statement, we must either be entitled to affirm it or be entitled to deny it.

Principle (3) is false, and future contingents are among many counterexamples to it. In order to be entitled to affirm “There will be a sea-battle tomorrow,” a speaker needs some ground for that assertion; he needs evidence that battle will be joined. Similarly, in order to be entitled to deny “There will be a sea-battle tomorrow,” he needs evidence that battle will not be joined. In a case where the admirals have not decided whether to fight, there may be no such evidence, so Aristotle’s case is a plausible—if rather sketchy—counterexample to (3). But it is not thereby a counterexample to (1), which is the correct formulation of Bivalence. The correct formulation does not mention affirmation or denial; it mentions only truth and falsity.

Can the Revised Argument be applied to show that “There will be a sea-battle tomorrow” is either true or false? I think it can. The relevant instance of excluded middle is in order: there is nothing amiss in the claim “Either there will be a sea-battle tomorrow or there won’t be.” Moreover, “There will be a sea-battle tomorrow” seems to have a determinate sense. As uttered on 1 January 2015, what it says is that a sea-battle will take place on 2 January 2015. If a sea-battle does take place on the latter date, the utterance is true; if no sea-battle takes place then, it is false. In order for the utterance to be true, it is not necessary that battle was inevitable on 1 January; and in order for the utterance to be false, it is not necessary that battle was then precluded. So the claim that the statement is either true or false carries no deterministic implications.

At least, this is so if we understand “true” and “false” “atemporally”—i.e. in a sense that does not permit significant tensing.⁶ It is this understanding of these words that (T^*) and (F^*) articulate—for the quantification in (T^*) and (F^*) is itself tenseless. In this atemporal sense, the statement’s truth does not imply that a battle was determined by 1 January, and its falsity does not imply that a battle was then precluded. When “true” and “false” are taken in this sense, Bivalence is consistent with the future’s being open.

What further muddies the waters here, and gives the present attack on Bivalence some spurious plausibility, is that some English speakers use “true” in a way that *does* permit significant tensing. They say things like “It is already true (now, on 1 January 2015) that there will be a total solar eclipse, visible from the Faroe Islands, on 15 March 2015”—meaning thereby that the occurrence of such an eclipse on that date is already settled, fixed, or determined. Now in *this* sense of “true,” the Revised Argument would be fallacious and its conclusion false. Whilst it is now determined that there either will or will not be a sea-battle tomorrow, it does not follow that it is

⁶ This paragraph and the next summarize the argument of von Wright (1984), to whom I refer the reader for further elaboration.

either now determined that there will be a battle tomorrow or now determined that there will not be. To the contrary: if the admirals have yet to decide whether to fight, that conclusion will be false. The Revised Argument fails because (T^*) does not capture this temporal sense of “true” (and (F^*) does not capture the corresponding sense of “false”). But the failure of the Argument, and the falsity of its conclusion, when “true” and “false” are understood temporally are irrelevant to the Argument and to the Principle in the sense articulated here, in which “true” and “false” are understood atemporally. When the Principle is understood in that latter way, future contingents pose no threat to it whatever.

In his essay “Truth,” Michael Dummett considers a statement that may be regarded as an infinitary version of Aristotle’s example of the sea-battle: “A city will never be built on this spot” (Dummett 1959: 16). Dummett takes a statement’s content to be given by the commitments that a speaker who affirms it thereby incurs (i.e. by the “requirements” that must be satisfied if his assertion is to have been correct (1959: 22)). He also holds that a speaker incurs a commitment only if there is some finite bound on the time by which that commitment will either have been fulfilled or not. Let v be an utterance of “A city will never be built on this spot,” made on day d at place π . Then the commitments of v are these: that there be no city at π on day $(d + 1)$; that there be no city at π on day $(d + 2)$; ... Or, using Prior’s quantifiers into sentence position:

$$\forall P \forall n [\text{Commit}(v, P) \leftrightarrow (P \leftrightarrow \text{there is no city at } \pi \text{ on day } d + n)].$$

Now if we think of content in these terms, it is plausible to count a statement as true if all its commitments are fulfilled:

$$(T^D) \quad \forall u (\text{True}(u) \leftrightarrow \forall P (\text{Commit}(u, P) \rightarrow P)).$$

Similarly, it is plausible to take a statement to be false if one of its commitments is not fulfilled:

$$(F^D) \quad \forall u (\text{False}(u) \leftrightarrow \exists P (\text{Commit}(u, P) \wedge \neg P)).$$

So in the present case we shall have

$$\text{True}(v) \leftrightarrow \forall n (\text{there is no city at } \pi \text{ on day } d + n)$$

and

$$\text{False}(v) \leftrightarrow \exists n (\text{there is a city at } \pi \text{ on day } d + n).$$

These specifications of v ’s truth- and falsity-conditions seem to be correct: Dummett’s requirement that commitments be met or unmet in a finite time does not stop (T^D) and (F^D) from delivering the desired results in this case. Indeed, given classical logic, we can

now prove that v is bivalent. A theorem of first-order classical logic is $\lceil \forall x \neg Fx \vee \exists x Fx \rceil$. In particular, then, where $\lceil Cn \rceil$ symbolizes $\lceil \text{There is a city at } \pi \text{ on day } (d + n) \rceil$ we have $\lceil \forall n \neg Cn \vee \exists n Cn \rceil$,⁷ which combines with our conditions for v 's truth and falsity to yield the conclusion that v is either true or false.

Dummett, however, insists that we cannot assert the bivalence of v . In part this reflects his rejection of classical logic, and I shall return to the validity of $\lceil \forall x \neg Fx \vee \exists x Fx \rceil$ in the next section. His claim that we cannot assert v 's bivalence also rests, though, upon a thesis about truth. There is, he writes, an "important feature of the concept of truth which is not expressed by the law 'It is true that p if and only if p ' and which we have so far left quite of account: that a statement is true only if there is something in the world in virtue of which it is true" (Dummett 1959: 14). The things in virtue of which a statement is true, he explains, are "the sort of fact we have been taught to regard as justifying us in asserting it" (1959: 16). Let us call such a fact a *ground* of the statement. In these terms, Dummett is making the following claim:

$(T^G) \quad \forall u (\text{True}(u) \rightarrow \text{a ground for } u \text{ obtains}).$

Dummett holds that there is a corresponding thesis about falsehood. Let us say that a fact is an *anti-ground* of a statement if its obtaining justifies us in denying the statement. Then he also maintains,

$(F^G) \quad \forall u (\text{False}(u) \rightarrow \text{an anti-ground for } u \text{ obtains}).$

(T^G) and (F^G) together entail

$(Biv^G) \quad \forall u ((\text{True}(u) \vee \text{False}(u)) \rightarrow (\text{a ground for } u \text{ obtains}) \vee (\text{an anti-ground for } u \text{ obtains}))$

and, in the case of v , we are not entitled to assert the consequent of the relevant instance of (Biv^G) :

We are entitled to say that a statement P must be either true or false, that there must be something in virtue of which either it is true or it is false, only when P is a statement of such a kind that we could in a finite time bring ourselves into a position in which we are justified either in asserting or in denying P ; that is, when P is an effectively decidable statement. This limitation is not trivial: there is an immense range of statements which, like "Jones was brave" [said of a man who died without facing danger], are concealed conditionals, or which, like "A city will never be built here," contain—explicitly or implicitly—an unlimited generality, and which therefore fail the test. (Dummett 1959: 16–17)

⁷ In fact, this particular instance of excluded middle may be justified otherwise than by appeal to classical logic; see fn. 8.

We should agree with Dummett that we may not be entitled to assert that either a ground or an anti-ground for ν obtains. A ground for ν is the sort of fact that justifies us in asserting “A city will never be built here.” We know what sort of facts these are: the place in question is too cold to support a city, or it is too hot, or it lacks water, etc. Let us suppose, then, that none of the features afflicts π . The locus of ν , in other words, is a place where a city *could* well be built. On that supposition, no ground for ν obtains. An anti-ground for ν is the sort of fact that justifies us in denying “A city will never be built here.” We know what sort of facts these are too: population growth makes the urbanization of the place inevitable, or plans to build a city are already laid, etc. Let us suppose that none of these facts obtains either. In that case, we shall not be entitled to say that either a ground or an anti-ground for ν obtains.

But does it follow that we cannot assert that ν is bivalent? I think not. Dummett maintains that (T^G) is a general constraint on the application of the predicate “true”: truth is an “epistemically constrained” notion. In “Truth,” however, he offers no defense of this claim and it is striking that (T^G) is not needed to derive the intuitively correct truth-conditions for ν . It is, then, open to a philosopher to defend Bivalence by maintaining that ν is a counterexample to (T^G) and hence to (Biv^G) . Indeed, ν is a particularly strong form of counterexample to (T^G) , for ν may be true even though a ground for ν *never* obtains. It might be that π always remains a place where a city *could* be built—so that no one is ever justified in asserting “A city will never be built here”—while, as a matter of fact, no one ever happens to build a city there—so that “A city will never be built here” is true. This is why it is helpful to place Dummett’s example alongside Aristotle’s. Today, on 1 January, we can assert that “There will be a sea-battle tomorrow” is either true or false. Because the admirals have not decided whether to fight, no one today can assert that the statement is true and no one today can assert that it is false. Tomorrow, though, some people will be able to assert either that the statement is true or that it is false; by 2 January, either a ground or an anti-ground will have obtained. From a perspective that rejects (T^G) , the proper moral of Dummett’s example is that even this eventual obtaining of either a ground or an anti-ground is inessential to a statement’s bivalence. We may assert that ν is either true or false even though we know that neither a ground nor an anti-ground for ν may ever obtain. Dummett’s case brings out something interesting, then, but it is not a counterexample to Bivalence.

15.4 THE CONTINUUM HYPOTHESIS: THE INTUITIONIST ANALYSIS

A second class of putative counterexamples is more problematic.

When Cantor hypothesized “There is no set strictly intermediate in cardinality between the integers and the real numbers (the continuum),” it seems he succeeded in expressing a thought. Many philosophers, however, resist the claim that this statement—Cantor’s celebrated Continuum Hypothesis (CH for short)—is bivalent. Gödel (1940)

showed that the truth of CH was consistent with the axioms of Zermelo-Fraenkel set theory including the Axiom of Choice (assuming that those axioms are themselves consistent). Cohen (1966) proved that its falsity (i.e. the truth of its negation) was also consistent with the same axioms (under the same assumption). These results do not refute the claim that CH is either true or false, but they do cast doubt upon it. *Contra* Dummett, a statement may be true even though there are no (and never will be) grounds for asserting it. But many philosophers believe that a true (false) statement must have a *basis*: if a statement is true (false), there must be something that makes it true (false), even if we can never find out whether that basis obtains. The results of Gödel and Cohen appear to show that, in the case of CH, there is no basis for its truth or falsity in the generally accepted axioms of set theory.

But, if CH is not bivalent, where does the Revised Argument go wrong when applied to it? The intuitionists hold that we are not entitled to assert that CH is bivalent; on their view, the fallacious step in the Revised Argument is the appeal to the classical logical law of excluded middle at step (5). They identify “the principle of excluded middle with the principle of the solvability of every mathematical problem” (Brouwer 1927: 491). Since no one is entitled to assert that the continuum problem is solvable, no one is entitled to assert line (5) of the Argument in the relevant case. As I now argue, though, the intuitionists do not properly capture the nature of mathematical doubts about the bivalence of statements like CH.

The semantic theory that Arend Heyting (1934) developed for the language of intuitionist mathematics reflects Brouwer’s identification of excluded middle with universal solvability. According to that theory,

the meaning of each [logical] constant is to be given by specifying, for any sentence in which that constant is the main operator, what is to count as a proof of that sentence, it being assumed that we already know what is to count as a proof of any of the constituents. (Dummett 2000: 8)

Specifically, Heyting stipulated that a proof of $\lceil A \vee B \rceil$ is anything that is a proof either of A or of B , and that a proof of $\lceil \neg A \rceil$ is a construction of which we can recognize that, when applied to any proof of A , it will yield a proof of a contradiction. A statement counts as intuitionistically valid if the semantic principles guarantee it to be provable no matter which atomic statements are provable. So a statement of the form $\lceil A \vee \neg A \rceil$ will be valid only if either A or $\lceil \neg A \rceil$ is provable, i.e. only if the problem of deciding A is solvable. On this semantics, to take $\lceil A \vee \neg A \rceil$ to be valid, no matter what mathematical statement A might be, would precisely be to postulate that every mathematical problem is solvable.

At first blush, the Heyting semantics seems to give us what we want. Since we cannot assert that the continuum problem is solvable, we cannot assert the relevant instance of excluded middle in line (5). So, in intuitionistic logic, the Revised Argument for the bivalence of the Continuum Hypothesis barely gets started.

There is a snag, though. Consider Goldbach’s Conjecture: “Every even number greater than two is the sum of two primes.” At the time of writing, no one has a proof or

a refutation of this conjecture, or a demonstration that a proof or refutation must exist. If the statement is false, a refutation will exist. For if the statement is false, there will be a counterexample to the Conjecture, and in that event it will be possible in principle to identify all the prime numbers less than the counterexample and then verify that no pair of them has the counterexample as its sum. However, there is at present no reason to assert that the Conjecture must have either a proof or a refutation. For all we know, there may be no counterexample to it, but at the same time no uniform reason why every even number greater than two is the sum of two primes, nor even a finite partitioning of those numbers with a uniform reason for each partition. Given the Heyting semantics, then, we are at present unable to assert $\lceil GC \vee \neg GC \rceil$ where GC states Goldbach's Conjecture. This is a problem, however, for almost every mathematician believes that we *can* assert $\lceil GC \vee \neg GC \rceil$, even in the face of the recognition that GC may not be decidable. For this reason, the intuitionist seems to give the wrong explanation of why CH is not bivalent.

What grounds the widespread belief that we can assert $\lceil GC \vee \neg GC \rceil$? In articulating these grounds, it helps to spell out some notions. Let us say that a property φ is *definite* (with respect to a domain) if each member of that domain is either φ or not φ . And let us say that a domain D is *determinate* when, for any property φ that is definite with respect to D , the following thesis holds when the range of the quantifiers is restricted to D :

(Det) $\forall x \varphi x \vee \exists x \neg \varphi x$.

The thought underlying (Det) is this. For a domain to be determinate is for it to be determinate which objects belong to it. Now where φ is definite with respect to D , each member of D is either φ or not φ . If it is also determinate which objects belong to D , then the question "Is every member of D a φ ?" must have one of its expected answers. If that answer is "yes," then $\forall x \varphi x$. If the answer is "no," then some member of D is not φ , so that $\exists x \neg \varphi x$. Either way, we have $\lceil \forall x \varphi x \vee \exists x \neg \varphi x \rceil$.

These notions enable us to spell out the basis for the widespread conviction that we are entitled to assert $\lceil GC \vee \neg GC \rceil$. Let the domain of quantification, D , be the even natural numbers greater than two; and let $\lceil Fn \rceil$ be the statement ' n is the sum of two primes'. Then Goldbach's Conjecture is expressed by the formula $\lceil \forall x Fx \rceil$. Now it is agreed on all hands (even by intuitionists operating under the Heyting semantics) that each even number greater than two is either the sum of two primes or it is not. In our terms, F is a definite property of the members of the domain D . Moreover, it is natural to hold that the domain of even numbers is determinate: i.e. it is determinate which objects are even numbers. For example, 2, 4, and 6 are even numbers, while 1, 3, and Julius Caesar are not. We can, then, assert $\lceil \forall x Fx \vee \exists x \neg Fx \rceil$. Even in intuitionistic logic, though, $\lceil \exists x \neg Fx \rceil$ entails $\lceil \neg \forall x Fx \rceil$, so $\lceil \forall x Fx \vee \exists x \neg Fx \rceil$ entails $\lceil \forall x Fx \vee \neg \forall x Fx \rceil$, i.e. $\lceil GC \vee \neg GC \rceil$.⁸

⁸ A parallel argument establishes the truth of $\lceil \forall n \neg Cn \vee \exists n Cn \rceil$, where $\lceil Cn \rceil$ means ' \lceil There is a city at place π on day $(d + n)\rceil$ '. (An empirical property of the natural numbers may still be definite in the specified sense.) The truth of $\lceil \forall n \neg Cn \vee \exists n Cn \rceil$ was the main premise of the positive argument given in 15.3 for the bivalence of Dummett's example v . That example may be regarded as an ingenious hybrid of Goldbach's Conjecture and Aristotle's Sea Battle.

The contentious premise of this argument is the claim that the domain of even numbers—and, with it, the domain of all natural numbers—is determinate. Whether these domains are determinate or not is a deep question in the philosophy of arithmetic that I cannot discuss here.⁹ Whatever the answer may be, though, our analysis shows where Brouwer's doubts about CH's bivalence diverge from those of mainstream mathematicians. The problem does not lie in the use of intuitionistic logic per se. In that logic, some instances of excluded middle may be asserted. Rather, the divergence stems from Brouwer's contention, which is embodied in the Heyting semantics, that $\lceil A \vee \neg A \rceil$ may be asserted only when A is decidable. For Brouwer, we cannot assert that CH is bivalent *because* we cannot assert that CH is decidable. For most mathematicians who have doubts about the bivalence of CH, though, that cannot be the proper explanation: they take themselves to be entitled to assert that Goldbach's Conjecture is bivalent even though they cannot assert that the Conjecture is decidable. If CH is not bivalent, its being so does not lie simply in its being undecidable.¹⁰

15.5 FAILURES OF BIVALENCE IN SET THEORY: A BETTER ANALYSIS

I think people are right to doubt the bivalence of many statements of set theory but I do not think that those doubts offer any support to Brouwer's intuitionism. Rather, I want to argue, the doubts are well-founded because many set-theoretic statements are indeterminate: they fail to be bivalent because they do not satisfy condition (D).

When a mathematician utters the sentence "There is no set strictly intermediate in cardinality between the integers and the real numbers (the continuum)," what is he saying? He is clearly saying something about sets, but what are sets? The word "set," as it comes from his mouth, is surely a theoretical term, and the theory that implicitly defines it—the theory that endows it with sense—can only be set theory. In uttering CH, then, the mathematician is saying something about the mathematical structure that is characterized by the axioms of set theory.

An account of the content of set-theoretic statements along these lines seems to me to be the only one that is remotely plausible.¹¹ There is, however, a problem with the

⁹ For an interesting discussion, see Field (1998).

¹⁰ Cf. Benacerraf and Putnam: "It is instructive to compare [and contrast] set theory with number theory. In number theory too there are statements that are neither provable nor refutable from the axioms of present-day mathematics. Intuitionists might <argue> that that this shows (not by itself, of course, but together with other considerations) that we do not have a clear notion of 'truth' in number theory, and that our notion of a 'totality of all integers' is not precise. Most mathematicians would reject this conclusion. Yet most mathematicians feel that the notion of an 'arbitrary set' is somewhat unclear" (1983: 19).

¹¹ For a persuasive elaboration of the account, see now Issacson (2011).

formulation I have given: there is no such thing as *the* mathematical structure (in the singular) that is characterized by the axioms of standard set theory. The point is that even second-order set theory, ZFC^2 , is not fully categorical. Its axioms leave (at least) the “height” of the set-theoretic universe undetermined, so two models of ZFC^2 need not be isomorphic. As Zermelo showed, the theory is only *quasi-categorical*: given any two models, one will be isomorphic to an initial sub-model of the other (Zermelo 1930). More exactly, each model of ZFC^2 has the form V_κ , where κ is a strongly inaccessible cardinal. (The hierarchy of sets V_α is defined as follows for each sort of ordinal α : $V_0 = \emptyset$; $V_{\alpha+1} = \wp(V_\alpha)$ for each successor ordinal $\alpha+1$; $V_\lambda = \bigcup_{\beta < \lambda} V_\beta$ for each limit ordinal λ . For present purposes, the cardinal number of a set A may be taken to be the least ordinal equinumerous with A .)

What are the consequences of this? One is that some theorems of first-order classical logic are hard to justify when the domain of quantification is understood to range over all sets. An example of such a theorem is (*Det*) from section 15.4:

$$\forall x \varphi x \vee \exists x \neg \varphi x.$$

I argued there that we are entitled to assert (*Det*) whenever the domain of quantification is determinate and φ is a definite property. Even when φ is definite, however, it is hard to see what entitles us to assert $[\forall x \varphi x \vee \exists x \neg \varphi x]$ when the variable “ x ” ranges over absolutely all sets. Certainly the justification given in §4 does not apply. That justification presumed that it was a determinate matter which objects are sets. But it is hard to see how this can be a determinate matter if (a) set theory is what determines which objects satisfy “set” whereas (b) set theory is not categorical.

Is the use of classical logic illegitimate, then, when reasoning about the totality of all sets? Not necessarily. For one might maintain that statements apparently involving quantification over all sets need to be tamed before we can do any systematic reasoning with them and that, once we tame them, it is legitimate to apply classical logic to them. The sort of “taming” I have in mind is already implicit in Zermelo’s writings. The very idea of non-categoricity presumes a multiplicity of non-isomorphic models of ZFC^2 ; but the domains of those models are themselves sets and, as such, are determinate. Thus “what appears as an ‘ultra-finite non- or super-set’ in one model is, in the succeeding model, a perfectly good, valid set with both a cardinal number and an ordinal type, and is itself a foundation stone for the construction of a new domain” (Zermelo 1930: 1233). Now when the domain of quantification is restricted to the members of some set, the laws of classical logic will apply. What is peculiar about statements apparently about all sets, on this view, is not that their variables range over a collection that is not a set. It is, rather, that the semantics of such statements must be understood in relation to a whole sequence of domains of quantification: “To the unbounded series of Cantor ordinals there corresponds a similarly unbounded double-series of essentially different set-theoretic models, in each of which the whole classical theory is expressed . . . This series reaches no true completion in its unrestricted advance, but possesses only relative

stopping-points, just those ‘boundary numbers’ [i.e. strongly inaccessible cardinals] which separate the higher model types from the lower” (ibid.).

This way of understanding statements involving apparent quantification over all sets may justify the use of classical logic in reasoning with them. It does so, however, at the price of rendering certain set-theoretic statements indeterminate. For we have here an example of the situation considered earlier (see p. 399), where the only available sense that can be attached to “ u says that P ” is “An interpretation of u under which it says that P is as legitimate as any other interpretation.” *Ex hypothesi*, the only legitimate interpretations of u are those which vindicate the classical logic by restricting the domain of quantification to the members of a standard model of ZFC². Under this hypothesis, what is our mathematician saying when he utters CH? Where κ_1 is the first strongly inaccessible cardinal, one maximally legitimate interpretation of his statement is that there is no set in V_{κ_1} that is strictly intermediate in cardinality between the integers and the reals. But, where κ_2 is the second strongly inaccessible cardinal, another maximally legitimate interpretation of his statement is that there is no set in V_{κ_2} that is strictly intermediate in cardinality between the integers and the reals. And so forth. In making a set-theoretic statement, a speaker presents himself as saying something about the universe of sets, but there is no such thing—no unique such thing, anyway. The various determinations of that universe generate various, but equally legitimate, interpretations of set-theoretic statements.

For this reason, many set-theoretic statements fail to satisfy condition (D). Let u be the statement “There is at least one strongly inaccessible cardinal,” let “ P ” abbreviate “There is at least one strongly inaccessible cardinal in V_{κ_1} ” and let “ Q ” abbreviate “There is at least one strongly inaccessible cardinal in V_{κ_2} .” An instance of (D) is

(**) Say(u, P) \wedge Say(u, Q) \rightarrow ($P \leftrightarrow Q$).

On the view we are considering, (**) is false. On that view, the only available interpretation that sustains a classical logic for “ P ,” “ Q ” etc. is one whereby “Say(u, P)” means “An interpretation of u under which it says that P is as legitimate as any other interpretation.” Under that interpretation, the antecedent of (**) is true. Because our set-theoretic axioms fail to single out a unique intended “universe” of sets, one maximally legitimate interpretation of u is that there is at least one strongly inaccessible cardinal in V_{κ_1} , while another is that there is at least one strongly inaccessible cardinal in V_{κ_2} . The consequent of (**), however, is false: P is false (κ_1 is not a member of V_{κ_1}) whereas Q is true.

The Revised Argument cannot be applied, then, to establish that an arbitrarily selected statement of set theory is bivalent. Since some set-theoretic statements do not satisfy (D), we cannot always affirm line (3) of the Argument. Indeed, if the theory that gives content to the notion of “set” is simply ZFC², many set-theoretic statements are *not* bivalent, u of the previous paragraph being a case in point. u “says” that P and “says” that Q , where Q but not P . Since u says that P , and not P , (T^*) entails that u is not true. Since u says that Q , and Q , (F^*) entails that u is not false. So u is neither true nor false.

Unlike Brouwer's analysis, the present ground for saying that certain set-theoretic statements are not bivalent is entirely consistent with holding that undecidable statements of number theory such as Goldbach's Conjecture *are* bivalent. The crucial difference is that there are theories which provide a fully categorical axiomatization of the natural-number structure: our comprehension of such a theory provides the basis for a conception of the domain of natural numbers as being determinate in the sense of section 15.4. Second-order Peano arithmetic (PA^2) is the most famous example of a categorical axiomatization of the natural-number structure. Some philosophers doubt if we really understand full second-order quantification over infinite domains, but even if we cannot grasp PA^2 , there are weaker theories—indeed, theories whose logic is weaker than full second-order logic—that characterize the natural numbers up to isomorphism. By the Löwenheim-Skolem Theorem, the logic of a categorical axiomatization of number theory cannot be first-order. But consider “ancestral logic,” a system in which first-order logic is supplemented with introduction and elimination rules for an operator $*$ which maps a relation to its ancestral (see Myhill 1952). (The $*$ -operator is counted as a logical constant.) In this system, we may formulate a theory, T , in which the axioms of first-order Peano arithmetic, PA^1 , are supplemented by the axiom $\lceil \forall x(x = 0 \vee S^*0x) \rceil$, where S is the relation of immediate succession. Unlike PA^1 , T characterizes the natural numbers up to isomorphism: since the new axiom requires each member of the domain to be only finitely many steps of immediate succession from zero, it excludes the non-standard numbers. However, the conceptual resources needed to make sense of T are much weaker than those needed to understand PA^2 . In particular, it is not necessary to make sense of the idea of an arbitrary set of natural numbers.

Many contemporary set theorists regard ZFC^2 as too weak a theory to characterize the notion of “set.” Most of the stronger theories that have been proposed do not threaten the above argument that not all set-theoretic statements are bivalent, for the theories in question remain only quasi-categorical. Vann McGee (1997), however, has proposed a new axiom that renders the resulting set theory categorical. McGee works in a system, ZFU^2 , in which the first-order quantifiers are permitted to range over *Urelemente*—objects that are not sets—as well as over sets. He proposes as a further set-theoretic axiom the postulate that these *Urelemente* form a set (call this axiom “McG”); he then proves that the theory $ZFU^2 + McG$ gives a categorical characterization of the pure sets—i.e. the sets that may be formed from the empty set using the standard set-theoretic operations. More precisely, McGee shows that there is a theorem of $ZFU^2 + McG$ which, under the intended interpretation of the language of ZFU^2 , says the following: any structure that could, given the axioms of $ZFU^2 + McG$, serve as the interpretation of “pure set” and “is a member of,” will be isomorphic to the intended structure of the pure sets.¹²

¹² That is, McGee shows that the theory of pure sets in $ZFU^2 + McG$ is “internally categorical” in the sense of Walmsley (2002: section 4).

McGee's result is striking, and one commentator has welcomed it as opening "the door to a structuralist account of set-theoretic truth on which every sentence of [the language of] pure set theory is assigned a determinate truth value" (Uzquiano 2002: 181). But have we reason to accept McGee's axiom? So far from its being an evident truth, I think it is highly doubtful.

Let us consider the notion of an ordinal, i.e. the order-type of a well-ordered set. Ordinals can be represented as sets. However, there are many equally good ways of so representing them. For the reasons expounded by Paul Benacerraf in "What Numbers Could Not Be" (Benacerraf 1965), this suggests forcibly that they are not themselves sets. If an ordinal were a set, there would have to be a fact as to which set it is. Given the multiplicity of equally good set-theoretic representations of the ordinals, it is hard to see what that fact could be.¹³ We shall do better to follow Christopher Menzel and treat ordinals as *properties* of well-ordered sets: an ordinal is something that isomorphic well-ordered sets have in common (Menzel 1986: 43).

On this view, an ordinal is an *Urelement*, so McG (together with Separation) implies that there is a set **On** of all ordinals. At this point, however, the Burali-Forti paradox looms (see Menzel 1986: 38). Certainly, that paradox does not need the assumption that ordinals are sets. All we need is that the putative set **On** is well-ordered by the relation $<$ of being strictly less than: since an ordinal just *is* the order-type of a well-ordered set, it follows immediately from this theorem that **On** possesses an ordinal, β . Now any ordinal is the order-type of the set of ordinals strictly less than itself. So β is also the order-type of the set A of all ordinals that are strictly less than β . Since A and **On** share an order-type, they must be isomorphic. It is easy to show, though, that no well-ordered set is isomorphic to any of its proper initial segments, so that A and **On** must be identical. Since **On** is the set of all ordinals, $\beta \in \mathbf{On}$, whence $\beta \in A$. But then, by the definition of A , it follows that $\beta < \beta$. Menzel (1986; 2014) has explored ways of avoiding outright contradiction here by restricting some of the other axioms of ZF. The fact remains, though, that McG's having this implication renders it highly doubtful. Nice as it might seem to have a categorical characterization of the universe of sets, we have reason not to accept McGee's theory.

15.6 KREISEL'S ARGUMENT FOR THE BIVALENCE OF CH

The argument of the previous section shows that some set-theoretic statements are not bivalent. Others are, though. Further confirmation of our analysis comes from the sense it makes of a famous argument purporting to show that CH, in particular, is bivalent.

¹³ For this application of Benacerraf's argument see Menzel (1986: 41–3). Of course, if the natural numbers *are* ordinals, this is not a new application of Benacerraf's argument but a reiteration of it. In "What Numbers Could Not Be," though, Benacerraf seemed to think of a natural number as something abstracted from its ordinal and cardinal applications.

The argument I have in mind was given by Georg Kreisel in section 1 of his paper “Informal Rigour and Completeness Proofs” (1967). Kreisel’s point of departure is Gödel’s well-known essay “What is Cantor’s Continuum Problem?” (1964). In that essay, Gödel was concerned to distinguish between the sort of independence that the Parallel Postulate has from the other axioms of Euclidean geometry, and the sort of independence that CH has from the axioms of ZFC. Gödel’s account of the difference is informal and somewhat vague; Kreisel aims to characterize it more precisely “in terms of higher-order consequence” (1967: 79). He duly begins section 1 by remarking that second-order logic has a good claim to be the implicit logic of the working mathematician:

The familiar classical structures (natural numbers with the successor relation, the continuum with a denumerable dense base, etc.) are definable [i.e. may be characterized up to isomorphism] by *second-order* axioms, as shown by Dedekind. Zermelo showed that his cumulative hierarchy up to ω or $\omega + \omega$, or $\omega + n$ (for fixed n) and other important ordinals is equally definable by second-order formulae. Whenever we have such a second-order definition there is associated a schema in first-order form (in the language considered): For instance, in Peano’s [induction] axiom

$$\forall P[\{P(0) \wedge \forall x [P(x) \rightarrow P(x+1)]\} \rightarrow \forall x P(x)]$$

one replaces the second-order quantifier $\forall P$ by a list of those P which are explicitly defined in ordinary first-order form (from $+$ and \times , for instance). A moment’s reflection shows that the evidence of the first-order axiom schema derives from the second-order schema: the difference is that when one puts down the first-order schema one is supposed to have convinced oneself that the specific formulae used (in particular, the logical operations) are well defined in any structure that one considers. (Kreisel 1967: 85–6)

This point enables Kreisel to put his finger on the relevant difference between the two cases of independence. Geometry, too, admits of a second-order axiomatization, and “*the parallel axiom is not even a second-order consequence*” of the other geometric axioms (Kreisel 1967: 88; Kreisel’s emphasis). So there really are different geometries, in some of which the Parallel Postulate holds, in others of which it does not. Consequently, it makes no sense to ask whether the Postulate is true or false *simpliciter*; all one can say is that it holds in some geometries and does not hold in others. For the question of its truth or falsity to be meaningful, we would need to give some extra-axiomatic specification of the meaning of the geometric theory’s primitive terms—for example by stipulating that a straight line is the path that a light-ray traverses *in vacuo*. Once that stipulation had been made, it would make sense to ask whether the Parallel Postulate is true or false, and the question would be settled by empirical investigation (our best theory of the path of light-rays says that the Postulate is false). *Per contra*, “CH is only independent of the *first-order* schema (associated with the axioms) of Zermelo-Fraenkel” (ibid.)—i.e. of the schema that is associated with the second-order Zermelo-Fraenkel axioms in the way in which the formula with “ P ” as a schematic letter is associated with the second-order induction axiom displayed above.

Kreisel surely identified an important point of difference between the independence of the Parallel Postulate from the other axioms of classical geometry and the independence of CH from the axioms of ZFC. But how, exactly, is that point supposed to establish that CH is bivalent? Here is the crucial passage:

Let Z be Zermelo's axiom with the axiom of infinity, and let CH be the (canonical) formulation of the continuum hypothesis in the following form: if C_ω is the collection of hereditarily finite sets without individuals, $C_{\omega+1} = C_\omega \cup \wp(C_\omega)$, $C_{\omega+2} = C_{\omega+1} \cup \wp(C_{\omega+1})$, CH states that

$$X \subset C_{\omega+1} \rightarrow (\|X\| \leq \|C_\omega\| \vee \|X\| = \|C_{\omega+1}\|),$$

which is expressed by means of quantifiers over $C_{\omega+2}$. As Zermelo pointed out, if we use the current set-theoretic definition $Z(x)$ of the cumulative hierarchy, in any model of Z , this formula Z defines a C_σ for a limit ordinal $\sigma > \omega$. Consequently we have

$$(Z \vdash_2 \text{CH}) \vee (Z \vdash_2 \neg \text{CH}) \quad (1967: 87-8).$$

Although the single turnstile " \vdash_2 " is standardly used to signify deducibility in axiomatic second-order logic, Kreisel does not (or should not) mean this: Weston (1976: 289-90) extends an argument of Tharp's to show that neither CH nor $\neg \text{CH}$ is a theorem of axiomatic ZFC². The proper conclusion of his argument is that either $Z \vdash_2 \text{CH}$ or $Z \vdash_2 \neg \text{CH}$, where " \vdash_2 " signifies the model-theoretic consequence relation of full second-order logic.¹⁴

Kreisel's argument is compressed, but it may be spelled out as follows. Let us begin by considering a theory, T , formulated in a second-order language L , and let us suppose that T is categorical—i.e. that all its models are isomorphic. Consider now two such models, M and N . A general theorem tells us that the same closed sentences of L are true in the two models. That is, for any closed sentence σ of L , $M \models \sigma$ if and only if $N \models \sigma$. The proof of this theorem is in essence straightforward. Since T is categorical, there is an isomorphism f from M to N . One then shows by induction on complexity that, for any expression e of L , the semantic value of e with respect to N is the result of applying f to the value of e with respect to M . This shows that if $N \models \sigma$ then $M \models \sigma$, for any closed

¹⁴ Kreisel gives his argument for Zermelo's original (1908) axiom system Z , which lacks the Axiom of Replacement, rather than for the more familiar system ZF , which includes it. (Kreisel clearly uses " Z " to stand for Zermelo's original system rather than for ZF itself. He observes (1967: 88) that while CH is determined by the second-order axioms of Z , Replacement is not; this would make no sense if " Z " meant ZF .) This renders his argument problematical, for Z^2 is not even quasi-categorical. Indeed, in the context of Z^2 , standard formulations of the axiom of infinity turn out to be non-equivalent, and some interpretations of $Z^2 + \text{Infinity}$ have non-well-founded models (see Uzquiano 1999). A charitable exposition of Kreisel's argument cuts through these complexities by taking it to apply to ZFC, not Z .

sentence σ of L . Since there is also an isomorphism from N to M , a parallel argument establishes the converse.

As we have seen, because ZFC^2 is only quasi-categorical, there are closed sentences in its language that are true in some of its models and false in others. Kreisel's insight, though, was to recognize that CH is not subject to this variation in truth-value between models. As he remarks in the last passage quoted from him, CH may be formulated using quantification over sets at levels $V_{\omega+2}$ and below. All the models of ZFC^2 , *when restricted to those levels*, are isomorphic, so the argument of the previous paragraph shows that, for any models M and N of ZFC^2 , $M \models_2 \text{CH}$ if and only if $N \models_2 \text{CH}$. The argument rests on particular features of CH; it turns on the fact that CH quantifies only over sets that lie so low in the set-theoretic hierarchy that every model of ZFC^2 will include them.

The argument as presented so far, however, does not take us all the way to Kreisel's conclusion. The quasi-categoricity of ZFC^2 combines with the particular features of CH to yield

- (1) For any models M and N of ZFC^2 , $M \models_2 \text{CH}$ if and only if $N \models_2 \text{CH}$.

Kreisel's eventual conclusion is the bivalence of CH:

- (B) CH is either true or false.

How are we supposed to move from (1) to (B)?

The natural reconstruction runs as follows. A familiar principle of classical model theory says that, for any model, any closed sentence is either true in the model or false in it. So, in particular,

- (2) For any model M of ZFC^2 , either $M \models_2 \text{CH}$ or $M \models_2 \neg\text{CH}$.

Together, (1) and (2) entail

- (3) Either $M \models_2 \text{CH}$ for any model M of ZFC^2 , or $M \models_2 \neg\text{CH}$ for any model M of ZFC^2 .

An abbreviated way of writing (3) is

- (4) Either $\text{ZFC}^2 \models_2 \text{CH}$ or $\text{ZFC}^2 \models_2 \neg\text{CH}$

which is what Kreisel wrote—or meant to write—in the last line quoted from him. But in order to get from (4) to (B) we need

- (5) If $\text{ZFC}^2 \models_2 \text{CH}$ then CH is true

and

- (6) If $\text{ZFC}^2 \models_2 \neg\text{CH}$ then CH is false.

What justifies (5) and (6)?

I think (5) and (6) may be justified as specializations, to the present case, of our general principles (T^*) and (F^*). According to (T^*), a statement will be counted as true if and only if, however it may legitimately be taken to say that things are, they are thus. As we have seen, the various models of ZFC^2 correspond to the various things that set-theoretical statements could legitimately be taken to be saying something about. So (T^*) yields

(5+) CH is true if and only if $ZFC^2 \vdash_2 CH$

from which (5) follows. Similarly, according to (F^*), a statement will be counted as false if and only if, however it may legitimately be taken to say that things are, they are not thus. As before, the models of ZFC^2 correspond to the ways set-theoretical statements could legitimately be taken to say that things are. So (F^*) yields

(6+) CH is false if and only if $ZFC^2 \vdash_2 \neg CH$

from which (6) follows.

In saying this, I am not endorsing Kreisel's argument. One doubtful step is line (2). Although (2) follows from a principle of classical model theory, the present application of that theory has been contested. Since the pure hereditarily finite sets are isomorphic to the set N of natural numbers, line (2) assumes that both $\wp(N)$ and $\wp(\wp(N))$ exist as well-defined sets, over which it is legitimate to quantify classically. The discussion of Kreisel's argument among philosophers of mathematics has focused on whether such quantification is legitimate (see especially Feferman 2009 and 2011, and Koellner 2010). All the same, our analysis confirms that his argument has the right "shape" to justify the bivalence of CH. In effect, Kreisel argues that CH is determinate (despite the general indeterminacy in the sense of "set") so that line (3) of the Revised Argument can be asserted.¹⁵

Even those set theorists who doubt the bivalence of some set-theoretic statements generally take classical logic for granted in doing set theory. Our analysis shows that their position is coherent. There can be failures of determinacy, and hence statements

¹⁵ Kreisel's claim that CH is bivalent is entirely consistent with the philosophical thesis (cf. the beginning of section 15.4) that a statement can be true (false) only if something *makes* it true (makes it false)—i.e. that truth and falsehood require a *basis*. If Kreisel's argument works at all, it shows that either $ZFC^2 \vdash_2 CH$ or $ZFC^2 \vdash_2 \neg CH$. So the basis of the truth or falsehood of CH (as the case may be) lies in the axioms of ZFC^2 . Since the consequence relation of full second-order logic cannot be completely axiomatized, it does not follow that either $ZFC^2 \vdash_2 CH$ or $ZFC^2 \vdash_2 \neg CH$, where " \vdash_2 " signifies deducibility in axiomatic second-order logic. Indeed, the Tharp-Weston result shows that this claim is false. If our knowledge of set-theory were confined to what we could deduce from ZFC^2 in axiomatic second-order logic, it would follow that neither CH nor $\neg CH$ has a *ground* in the sense of section 15.3 above. That is to say: we shall never be entitled to assert CH or to deny it. However, that is no threat to CH's bivalence, for we found reason to reject Dummett's claim that a statement can be true only if it has a ground and false only if it has an anti-ground. The case of CH, then, is one where it is important to distinguish between *basis* and *ground*.

that are neither true nor false, without any deviations from classical logic. It is noteworthy, however, that without the Principle of Bivalence, one cannot justify the classical logical laws on the basis of the specifications of the meanings of the connectives in the familiar classical truth-tables, even given a classical metalogic. Thus the truth-tables for “ \vee ” and “ \neg ” entail that $\lceil A \vee \neg A \rceil$ is true if A is either true or false, but we need Bivalence to infer from that that any instance of $\lceil A \vee \neg A \rceil$ is true. This need not disturb a set-theorist who is wedded to classical logic. Even if called upon to justify his adherence to that logic, there are alternative justifications—indeed, alternative semantic justifications—that he can pray in aid.¹⁶ Given a weak metalogic, Bivalence is sufficient to establish the validity of the classical logical laws, but it is far from being necessary.

15.7 THE CHALLENGE OF VAGUE STATEMENTS

Another class of putative counterexamples to Bivalence comprises statements in which a vague predicate attaches to a borderline case of the property that it signifies. Let us call such a statement *vague*. Imagine a hundred transparent tubes of paint, a_1, \dots, a_{100} , that steadily and almost imperceptibly change from clearly red to clearly orange as the subscript increases. Then consider the corresponding sequence A_1, \dots, A_{100} of statements, in which the statement A_n says that the tube a_n is red. Bivalence tells us that each of these statements is either true or false. Now if a statement attributing redness to an object a is true, and if b is redder than a , then a statement attributing redness to b will also be true. So Bivalence implies that there is a number N such that all of the statements A_1, \dots, A_N in our sequence are true and all of the statements A_{N+1}, \dots, A_{100} are false. It implies, in other words, that there is a cut-off point in the sequence at which the statements switch from being true to being false. But that in turn implies that there is a cut-off point at which the tubes switch from being red to being not red—a grossly implausible conclusion. Yet the only premises needed to reach that conclusion were Bivalence and a highly intuitive assumption about how the predicate “red” is to be applied.

One way of resisting the claim that vague statements are bivalent runs parallel to the treatment of set-theoretic statements recommended in sections 15.5 and 15.6. On this approach, only sentences with completely precise conditions of truth may replace the propositional variables “ P ” “ Q ” etc., so the applications of classical logic in the Revised Argument are unexceptionable. When the variables are interpreted in this way, vague statements will not be determinate in sense: a vague statement may be interpreted equally legitimately as saying that P and as saying that Q , where “ P ” and “ Q ” are inequivalent. Some of these precise specifications of its sense come out true and others

¹⁶ See Rumfitt (2015: ch. 9) for one of them.

come out false, so (T^*) and (F^*) entail that the statement is neither true nor false. This conclusion is consistent with the soundness of the Revised Argument, whose conclusion says only that every *determinate* statement is either true or false.

This approach yields a supervaluational treatment of vagueness, and so exhibits what is at once the great strength and the great weakness of any such treatment, its concern to preserve classical logic.¹⁷ Just as set-theorists use classical logic to construct their proofs, even when they doubt or deny the bivalence of some of the statements in those proofs, so ordinary thinkers apply the classical laws to vague statements, even as they doubt or deny *their* bivalence. All the same, there are grounds for querying the application of classical logic to vague statements that have no counterpart in the case of set theory. For by applying classical logic to vague statements we can derive apparently absurd conclusions, without making any appeal to Bivalence.

This is shown, of course, by versions of the Sorites Paradox. Let us revert to our hundred tubes of paint and the statements that attribute redness to each of them. The ground for denying that all of those statements are bivalent was that Bivalence entails the existence of a sharp boundary to the red tubes, and that consequence is absurd. But if that consequence is absurd, we ought to be able to assert its negation. Now there will be a sharp boundary to the red tubes if some tube in the sequence is red and its successor is not; that is if $[A_n \wedge \neg A_{n+1}]$ is true for some n . So the claim that there is a sharp boundary may be formulated as a long disjunction $[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})]$, and the claim that there is none as its negation. That is, we ought to be able to assert the following:

$$(1) \neg[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})].$$

In the situation described, though, we are also given that a_1 is red and that a_{100} is not red. So it seems that we ought also to be able to assert

$$(2) A_1$$

and

$$(3) \neg A_{100}.$$

Now if we suppose that a_{99} is red, i.e. if we suppose

$$(4) A_{99}$$

then the rule of conjunction-introduction applied to (3) and (4) would yield

$$(5) A_{99} \wedge \neg A_{100},$$

¹⁷ See Fine (1975) for the classic exposition and Varzi (2007) for a recent survey of the logical issues.

which, after ninety-nine applications of \vee -introduction, yields

$$(6) (A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})$$

which directly contradicts (1). Given (1) and (3), then, supposition (4) stands refuted, so by *reductio* we may assert

$$(7) \neg A_{99}.$$

By repeating this inferential sub-routine a further ninety-eight times, we reach

$$(8) \neg A_1$$

which contradicts (2). This, then, is the initial paradox. We have some reason to accept the trio of postulates (1), (2), and (3), but we also have an apparently valid deduction showing that the trio is inconsistent. It may be noted that the form of *reductio* that is applied in reaching line (7)—and that is re-applied at the corresponding later steps—is acceptable to an intuitionist. So the trio comprising (1), (2), and (3) is inconsistent in intuitionistic logic as well as in classical logic.

How should we react to this apparent demonstration of inconsistency? Since the case is one in which (2) and (3) are clearly true, it seems that we must take it as showing that (1) is false. In other words, we would appear to be entitled—indeed, compelled if the question of (1)’s truth arises—to make a further application of the relevant form of *reductio* and infer the negation of (1), namely,

$$(9) \neg \neg [(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})].$$

In classical logic, however, (9) is equivalent to

$$(10) (A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100}).$$

This, however, seems to land us in a yet more acute paradox, which Crispin Wright has called the *paradox of sharp boundaries* (see Wright 2007). For formula (10) says that at some point in the sequence a red tube is immediately followed by a non-red tube, and this seems to ascribe a sharp boundary to the red tubes. Wright calls this conclusion “unpalatable” and it does indeed seem to be something we are reluctant to accept. Given classical logic, however, it follows from the premises (2) and (3). That is, given only the premises “Tube a_1 is red” and “Tube a_{100} is not red,” classical logic yields a conclusion which seems to say that there is a sharp boundary to the red tubes in the sequence a_1, \dots, a_{100} . Little wonder that many philosophers have taken the Sorites to cast doubt on whether classical logic can be applied to deductions involving vague predicates.

Wright (2007) takes this form of the paradox as a ground for switching from classical to intuitionist logic. That logic validates all the steps up to line (9), but it does not validate

the elimination of double negation needed to reach (10). Wright's response, though, is not founded on any semantic analysis of vague expressions; as Dummett (2007) noted, the Heyting semantics certainly will not do.¹⁸ Because the choice of logic is in play, it would help to have a semantic model for vague statements against which intuitions about the validity of arguments involving them can be tested. But what might that model be?

15.8 THE SEMANTICS AND LOGIC OF VAGUE STATEMENTS

What is characteristic of vague expressions? An attractive general account was put forward by C. S. Peirce. In the entry "Vague (in logic)" that he contributed to Baldwin's *Dictionary of Philosophy and Psychology*, Peirce explained the word "vague" as meaning:

Indeterminate in intention.

A proposition is vague when there are possible states of things concerning which it is intrinsically uncertain whether, had they been contemplated by the speaker, he would have regarded them as excluded or allowed by the proposition. By intrinsically uncertain we mean not uncertain in consequence of any ignorance of the interpreter, but because the speaker's habits of language were indeterminate; so that one day he would regard the proposition as excluding, another as admitting, those states of things. Yet this must be understood to have reference to what might be *deduced* from a perfect knowledge of his state of mind; for it is precisely because those questions never did, or did not frequently, present themselves that his habit remained indeterminate. (Baldwin 1901–2, vol. 2: 748)

In remarking that a statement like "This is red" might exclude different states of affairs from day to day, Peirce points to an adaptability that vague terms confer upon a language. Precisely because the accepted sense of a vague statement is indeterminate, a speaker is free to render it more determinate by laying it down, as it might be, that the statement is to be taken as excluding certain possibilities whose status as excluded or allowed had hitherto been left unsettled. These determinations resemble decisions more closely than they resemble discoveries, so that indeterminacy of sense goes hand in hand with a form of semantic "freedom" whose importance for the analysis of vagueness Mark Sainsbury has stressed. The word "red" is associated with no boundary between the red things and the rest. But

it can be permissible to draw a line even where it is not mandatory to do so. No one can criticize an art materials shop for organizing its tubes of paint on various shelves,

¹⁸ See, though, Rumfitt (2018) for a semantic theory which, if accepted, would vindicate Wright's choice of intuitionistic logic as providing the standards for assessing arguments involving vague predicates.

including one labelled “red” and another “orange,” even though there is a barely detectable, or perhaps even in normal circumstances undetectable, difference between the reddest paint on the shelf marked “orange” and the most orange paint on the shelf marked “red.” Hence one can consistently combine the following: *red* draws no boundaries, that is, there is no adjacent pair in the series of tubes of paint such that the nature of the concept, together with the colour of the tube, requires one to apply *red* to one member of the pair but withhold it from the other; yet one can draw a boundary to the reds, that is, one may behave consistently with the nature of the concept in drawing a line between adjacent pairs.¹⁹

This seems right, as far as it goes, but I would add two points. First, many determinations of the sense of vague terms will be partial: even after the determination has been made, there will be some possible states of affairs of which it remains unsettled whether the statement is understood to exclude them. Sainsbury’s paint shop owner may determine the senses of “red” and “orange” so that precisely one of these predicates applies to every tube of paint in his shop. But there may be other tubes, redder than the reddest tube on his “orange” shelf but more orange than the most orange tube on his “red” shelf, whose classification remains undetermined even after he has settled how all the tubes currently in his shop are to be classified.

Second, in further determining the sense of a vague term, we often settle the truth-value of a complex statement without settling the truth-values of the components. Determination need not proceed from atoms to molecules. In particular, this holds for disjunctive statements: we often settle that a disjunction is true without settling which disjunct is true. A simple example of this will be familiar to readers who have been examiners in British universities. Undergraduate degrees in Britain are classified, and vague principles relate the candidate’s average numerical mark to his or her eventual class of degree. Thus the examiners’ deliberations are directed towards making the vague senses of “first-class,” “upper second-class,” etc. sufficiently determinate (in the given context) that one (and only one) of these predicates applies to each of that year’s finalists. Now suppose that there are two finals candidates, *A* and *B*, whose performance is borderline first-class. Suppose, in particular, that *A*’s average numerical mark is 69.25 while *B*’s is 69.2. Eventually, the examination board will have to determine the sense of the predicates of classification sufficiently precisely for each atom—in this case, each statement that ascribes a class to a candidate—to be either true or false. But there may be stages in the determination where it is settled that a disjunction is true, but where it is left open which disjunct is true. Thus, having re-read the scripts of *A* and *B*, the examining board may decide that their work is so similar in overall quality that the difference in their average mark reflects no real difference in the standard attained, so that the two candidates must be classified alike. Moreover, it may well reach this decision before settling which degrees *A* and *B* are to receive. This decision, then, amounts to determining the senses of “first-class” and “upper second-class” so as to verify the disjunction

¹⁹ Sainsbury (1990: 259–60). Here and in the next quotation, I have written “orange” where he writes “yellow.”

“Either *A* and *B* both receive first-class degrees, or *A* and *B* both receive upper second-class degrees” in advance of determining which disjunct is true. Examples to the same effect could easily be multiplied.

This suggests that we might model the meanings of vague statements by reference to a space of *partial determinations of sense*, or *pds*’s for short; a *pds* will partially determine the senses of all the vague expressions in the relevant language. When a given *pds* renders a statement true, I shall say that it *verifies* the statement. As in the example of the examiners’ classification, we allow that a *pds* may verify a complex statement even though it does not verify any of its parts. With the notion of verification in play, we may lay down the condition for consequence for a language containing vague terms: *B* follows from A_1, \dots, A_n just in case any *pds* that verifies all of A_1, \dots, A_n also verifies *B*. This account takes as given the extra-linguistic circumstances, which of course help to determine whether a given *pds* verifies a given statement. For the present, though, I shall suppress the way those circumstances vary in order to focus on the particular bearing that vagueness has on consequence.

A *pds* verifies a statement when it renders it true. In a similar spirit, we say that a *pds* falsifies a statement when it renders it false. It is natural to use the notion of falsification alongside that of verification in analyzing cases of vagueness. The customary sense of the word “red” is partial: it is indeterminate or “intrinsically uncertain” (in Peirce’s phrase) whether the term applies to our tube a_{50} . All the same, a grasp of that customary sense enables a speaker *both* to assert “Pillar boxes are red” and to deny “Oranges are red.” That is to say: the customary sense of “red” is marked out both by its verifying the first statement and by its falsifying the second. Because the customary sense is partial, we cannot infer that a statement is falsified from the premise that it is not verified. In order to characterize a *pds* adequately, we need to say both which statements it verifies and which it falsifies.

How are we to do this? Let us first define a dyadic relation \perp of *incompatibility* between *pds*’s as follows:

$x \perp y$ if and only if there is some possible statement *A* such that *x* falsifies *A* and *y* verifies *A*.

I postulate that incompatibility is (1) irreflexive and (2) symmetric. Postulate (1) says that no *pds* both verifies and falsifies a single statement. It amounts, in other words, to a requirement that every *pds* avoid the incoherence of entitling a speaker who uses terms in accordance with it simultaneously to assert and deny a given statement. As for postulate (2), let us assume that a speaker who denies *A* is committed to asserting *A*’s negation, and that a speaker who asserts *A* is committed to denying *A*’s negation. Given these assumptions, if *x* falsifies *A* and *y* verifies *A*, then *y* falsifies ‘not *A*’ and *x* verifies ‘not *A*’. That is, given these assumptions, if $x \perp y$ then $y \perp x$. Postulate (2), then, is sustained by the assumption that any possible statement has a negation with the following properties: any speaker who denies the original statement is committed to asserting the negation; and any speaker who asserts the original statement is committed to denying the negation. These properties are widely accepted attributes of negation.

This suggests that the space of pds's is an *orthoframe*—i.e a structure $\langle X, \perp \rangle$ comprising a non-empty set X of pds's, and a symmetric and irreflexive relation \perp between members of X . I use the notation $|A|$ to signify the set of pds's which verify the statement A . It is then straightforward to define the notion of falsification. Let us write $x \perp Y$ to mean that $x \perp y$ for every y belonging to a set Y , and consider the relation $x \perp |A|$. This will obtain when x is incompatible with every pds that verifies A ; i.e. when for *every* verifier y of A , x falsifies a statement which y verifies. This condition is met only when x falsifies A . Conversely, if x falsifies A , x will certainly be incompatible with any verifier of A . Given the meaning we have attached to \perp , then, “ x falsifies A ” is equivalent to “ $x \perp |A|$.” Although this obviates the need for separate semantic axioms concerning the conditions in which complex statements are falsified, the definition of the \perp -relation still presupposes a grip on the notion of denying a statement that does not reduce to asserting its negation.

Indeed, with the notion of falsification in play, we can lay down the semantic axiom for negation. A partial determination of sense will verify A 's negation precisely when it falsifies A itself:

$$(N) \quad |\neg A| = |A|^\perp.$$

(Where U is any set, I write U^\perp for $\{x: x \perp y \text{ for all } y \in U\}$.) The axiom for conjunction is also straightforward. A partial determination of sense will verify a conjunction just in case it verifies both conjuncts, so again we have

$$(C) \quad |A \wedge B| = |A| \cap |B|.$$

However, the analogous axiom for disjunction would be wrong. That axiom would be

$$(D1) \quad |A \vee B| = |A| \cup |B|,$$

but the case of the examination candidates is a counterexample to $(D1)$. We there had a partial determination of sense that verified a disjunction without verifying either disjunct.

How can we find a correct axiom for disjunction? The key is to attend to the topological properties of those sets that comprise a statement's verifiers. When Y and Z are subsets of X in an orthoframe $\langle X, \perp \rangle$, let us say that Y is \perp -closed in Z if the following condition is met:

$$\text{For all } x \in Z, x \notin Y \text{ only if there exists } y \in Z \text{ such that } y \perp x \text{ and } y \in Y.$$

I now argue that whenever sets Y and Z are semantic values of statements (i.e. whenever there are statements A and B such that Y is $|A|$ and Z is $|B|$), the following condition is met:

$$(\perp\text{-closure}) \quad \text{If } Y \subseteq Z, \text{ then } Y \text{ is } \perp\text{-closed in } Z.$$

I shall give the argument for particular choices of A and B ; it will be clear how it generalizes.

Let us take A to be the statement “Tube a_{n+1} is red” and B to be the statement “Tube a_n is red,” and let $Y = \{ \text{“Tube } a_{n+1} \text{ is red”} \}$ and $Z = \{ \text{“Tube } a_n \text{ is red”} \}$. Since tube a_{n+1} is less red than the tube a_n , any determination of the sense of “red” that renders A true also renders B true, so $Y \subseteq Z$. In order to establish \perp -closure, then, we need to show that Y is \perp -closed in Z .

What is it for Y to be \perp -closed in Z ? It comes to this. Consider an arbitrary pds, x , that verifies the statement “Tube a_n is red.” We need to show that if x does not verify “Tube a_{n+1} is red” then there is a pds, y , compatible with x , that verifies “Tube a_n is red” but falsifies “Tube a_{n+1} is red.” There are two cases to consider. If x falsifies “Tube a_{n+1} is red,” then x itself is such a y , for any pds is compatible with itself. If x does not falsify “Tube a_{n+1} is red,” then x leaves that statement undetermined—neither verified nor falsified. Now if a pds leaves a statement undetermined, there will be a further determination of it that renders it false. Such a further determination of x will of course verify “Tube a_n is red” (as x does) and will be compatible with x , so this further determination serves as the desired y . Either way, then, such a pds y exists, so the condition for Y to be \perp -closed in Z is met.

Although I have given the argument only for a particular choice of statements, it generalizes. The crux of the argument is that, if a pds leaves a statement undetermined, there will be a further determination of that pds that renders the statement false. I claim that this is a generally acceptable principle. Indeed, it is really just a manifestation of the “semantic freedom” that Sainsbury identified as being a mark of a vague term such as “red” and that follows from Peirce’s conception of vagueness as indeterminacy of sense. To quote Sainsbury’s formula once more, while “red” itself draws no boundary, “yet one can draw a boundary to the reds, i.e. one may behave consistently with the nature of the concept in drawing a line between adjacent pairs.” On the Peircean view, such freedom is a mark of all vague concepts. If that is right, the argument for \perp -closure will generalize to all vague statements.

We call a set \perp -closed (*simpliciter*) if it is \perp -closed in the whole space of pds’s, X . Since the verifiers of any statement form a subset of X , the verifiers of each statement must be \perp -closed (*simpliciter*). If sets Y and Z are \perp -closed, so are $Y \cap Z$ and Y^\perp . So, for complex statements built up using \wedge and \neg alone, the verifiers of every statement will be \perp -closed so long as the verifiers of each atom form a \perp -closed set.

This suggests the following definitions. Let us call the triple $\langle X, \perp, \xi \rangle$ a *vagueness frame* if $\langle X, \perp \rangle$ is an orthoframe and ξ is a non-empty collection of \perp -closed subsets of X such that (1) ξ is closed under set intersection and the orthocomplementation operation $^\perp$, and (2) if $Y, Z \in \xi$, then $Y \subseteq Z$ only if Y is \perp -closed in Z . We then call the quadruple $\langle X, \perp, \xi, V \rangle$ a *vagueness model* if V is a function which assigns a member of ξ to be the verifiers of each atomic statement of the relevant language and which respects axioms (C) and (N) above. Since $Y \cap Z$ and Y^\perp are \perp -closed whenever Y and Z are, V will assign members of ξ to complex statements built up using conjunction and negation. We can then say that a conclusion follows logically from some premises if every vagueness model that verifies all the premises also verifies the conclusion.

Our account of the constraints on the assignment function V is incomplete, for we have as yet said nothing about disjunction. Assuming contraposition in the metalogic, the condition for Y to be \perp -closed is that $x \in Y$ whenever $\forall y (y \perp Y \rightarrow y \perp x)$. That is to

say, it is equivalent to the condition $Y^{\perp\perp} \subseteq Y$. By the symmetry of \perp , any set is such that $Y \subseteq Y^{\perp\perp}$, so Y is \perp -closed if and only if $Y^{\perp\perp} = Y$.

This analogy points the way toward the correct semantic axiom for disjunction. We surely want the entailments from A to $\lceil A \vee B \rceil$ and from B to $\lceil A \vee B \rceil$ to come out valid, so $\lceil A \vee B \rceil$ must include $(\lceil A \rceil \cup \lceil B \rceil)$. Since the verifiers of any statement are \perp -closed, it is natural to take $\lceil A \vee B \rceil$ to be the smallest \perp -closed set that includes $(\lceil A \rceil \cup \lceil B \rceil)$. Thus we reach

$$(D) \quad \lceil A \vee B \rceil = (\lceil A \rceil \cup \lceil B \rceil)^{\perp\perp}.$$

Since $(Y \cup Z)^{\perp\perp} = (Y^{\perp} \cap Z^{\perp})^{\perp}$, (D) is equivalent to $\lceil A \vee B \rceil = (\lceil A \rceil^{\perp} \cap \lceil B \rceil^{\perp})^{\perp}$.

What logic does the proposed semantic theory validate? My definitions of a vagueness frame and of a vagueness model are exactly parallel to Goldblatt's definitions of a quantum frame and a quantum model, and it is quantum logic that turns out to be sound and complete with respect to the recommended notion of consequence (see Goldblatt 1974: 32–4). Quantum logic is usually formalized as a “binary” system, in which consequence is taken to relate a single statement serving as premise to a single conclusion, but it may also be formalized as characterizing a relation $X: A$ between a finite set of premises X and a single conclusion A . That formalization includes the usual structural rules of Reflexivity, Dilution and Cut, and the following rules for the operators:

\wedge -intro If $X: A$ and $Y: B$ then $X, Y: A \wedge B$

\wedge -elim If $X, A, B: C$ then $X, A \wedge B: C$

\vee -intro If $X: A$ then $X: A \vee B$ and $X: B$ then $X: A \vee B$

\vee -elim If $A: C$ and $B: C$ then $A \vee B: C$

DN $A: \neg\neg A$ and $\neg\neg A: A$

Con If $A: B$ then $\neg B: \neg A$

Red If $X: A$ then $X, \neg A: \perp$

DS If $B: A$ then $A \wedge (\neg A \vee B): B$.

Quantum logic does not contain the unrestricted version of \vee -elimination (with side-premises): if $X, A: C$ and $Y, B: C$ then $X, Y, A \vee B: C$. The form of Disjunctive Syllogism that I have called DS, and which is a “mixed rule” in that involves all three operators, partly compensates for the loss of deductive power that the restriction on \vee -elimination engenders. (One may prove that these rules are sound and complete with respect to the recommended account of consequence by adapting the completeness proof for quantum logic in Goldblatt 1974: section 6.)

15.9 THE PROBLEMS RESOLVED

How does this help to diagnose the flaw in the Sorites argument of section 15.7?

Where A_n is the statement “Tube a_n is red,” the first phase of that argument was a demonstration that a trio of plausible premises was inconsistent. We formulated the first of these premises—which says that the red tubes have no sharp boundary—as the negation of a disjunction, *viz.* $\lceil \neg[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})] \rceil$. But since all of De Morgan’s Laws are valid in quantum logic, I shall now take the equivalent formula below to be the canonical formulation of premise (1):

$$(1) \quad \neg(A_1 \wedge \neg A_2) \wedge \dots \wedge \neg(A_{99} \wedge \neg A_{100}).$$

As before, premises (2) and (3) of the Sorites argument are A_1 and $\neg A_{100}$. Let us call premise (1), “ B ”

The first part of our soritical deduction was directed towards showing that premises (1), (2), and (3) are an inconsistent trio—in other words, that $B, A_1, \neg A_{100} : \perp$. Our rules enable us to derive this sequent as follows. From Reflexivity and \wedge -elim we have $B : \neg(A_1 \wedge \neg A_2)$, which gives $B : \neg A_1 \vee A_2$ by De Morgan and DN. Since $A_2 : A_1$, the rule DS yields $A_1, \neg A_1 \vee A_2 : A_2$ whence $B, A_1 : A_2$ by Cut. A similar deduction shows that $B, A_2 : A_3$, whence $B, A_1 : A_3$ by Cut. By repeating this inferential sub-routine a further ninety-eight times, we eventually reach $B, A_1 : A_{100}$, whence $B, A_1, \neg A_{100} : \perp$ by Red. Our analysis confirms, then, that premises (1), (2), and (3) are an inconsistent trio.

What about the next stage of the Sorites argument, which moves from the two apparently incontrovertible premises—(2) and (3)—to the negation of premise (1)? Our rules validate this stage of the argument too. We already have $A_1, B : A_{100}$, which yields $A_1 \wedge B : A_{100}$ by \wedge -elim. This gives $\neg A_{100} : \neg(A_1 \wedge B)$ by Con, whence $\neg A_{100} : \neg A_1 \vee \neg B$ by De Morgan. We also have $\neg B : A_1$. The formula $\lceil \neg B \rceil$ can only be rendered true if some tube in the sequence is red; since a_1 is the reddest tube in the sequence, it must be counted red if any tube is, so any pds that verifies $\lceil \neg B \rceil$ will verify A_1 . Accordingly DS yields $A_1, \neg A_1 \vee \neg B : \neg B$. Cut then gives $A_1, \neg A_{100} : \neg B$. This validates the second stage of the original argument: A_1 and $\neg A_{100}$ are its premises (2) and (3), and $\lceil \neg B \rceil$ is the negation of premise (1).

At this stage, it may seem as though we have made no progress in escaping from the Sorites paradox. Premises (2) and (3) are highly plausible: if we can be sure of the truth-values of any statements involving vague terms, we can be sure that our paradigm red tube a_1 is red and that our paradigm orange tube a_{100} is not. So it seems that we must accept the conclusion $\lceil \neg B \rceil$. But B is equivalent to the long negated disjunction $\lceil \neg[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})] \rceil$, and hence $\lceil \neg B \rceil$ is equivalent to the doubly negated disjunction $\lceil \neg \neg[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})] \rceil$. In quantum logic double negations are eliminable, so our conclusion $\lceil \neg B \rceil$ is equivalent to the long disjunction

$$(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100}),$$

a formula I shall label C . C is simply Wright's "unpalatable" conclusion, unpalatable because it seems to say that there is a cut-off point somewhere in the sequence. Our logic, then, has taken us from two apparently certain premises—viz., A_1 and $\neg A_{100}$ —to an apparently unacceptable conclusion, viz. C . So it may seem that no substantial progress has been made in resolving the paradox.

In fact, though, we have made progress. For, when C is understood in the way our semantic theory requires, it is in truth entirely palatable. C appears to be unpalatable because it seems to assert the existence of a sharp boundary between the red and the non-red tubes. We imagine that C can be true only if one of its component disjuncts is true—say $\lceil A_{50} \wedge \neg A_{51} \rceil$. In turn, that disjunct can be true only if there is a sharp boundary between the red members of the sequence— a_1, \dots, a_{50} —and the non-red members: a_{51}, \dots, a_{100} . Given our semantics, though, this assumption about the truth-conditions of disjunctions is not correct. The semantic axiom for "or" does not say that a disjunction is true just when one of its disjuncts is true. Rather, it says that a partial determination of sense verifies a disjunction if it belongs to the *closure* of the union of the verifiers of the disjuncts, where closure is double orthocomplementation. It is therefore entirely possible for a pds to verify a disjunction without verifying either disjunct and this makes it possible for a disjunction to be true without either disjunct's being true.

Indeed, when we think what the present closure operator means, we can see how C may be true even though the concept *red* lacks a sharp boundary. Our semantic axiom for disjunction says that $|A \vee B| = (|A| \cup |B|)^{\perp\perp}$. A pds belongs to $U^{\perp\perp}$ if it is incompatible with every pds that is itself incompatible with all the members of U . Hence a pds verifies C if it is incompatible with every pds that falsifies the claim that there is a cut-off point in the sequence of red tubes. Now a pds falsifies that claim only when *either* the entire sequence is red *or* the entire sequence is not red. Given premises (2) and (3)—that tube a_1 is red and that tube a_{100} is not red—neither of these possibilities obtains. So the customary current sense of the word "red," partial as it is, is indeed incompatible with every pds in the orthocomplement of the union

$$|A_1 \wedge \neg A_2| \cup \dots \cup |A_{99} \wedge \neg A_{100}|$$

and hence belongs to $[(A_1 \wedge \neg A_2) \vee \dots \vee (A_{99} \wedge \neg A_{100})]$. It belongs to that set, and hence verifies C , because it verifies " a_1 is red" while falsifying " a_{100} is red." But because that pds verifies no particular disjunct $\lceil A_n \wedge \neg A_{n+1} \rceil$, the truth of C carries no commitment to the existence of a sharp cut-off in the sequence of red tubes. As one might put it: C says that there is a shift from red to non-red tubes *somewhere* in the sequence, but there need be no determinate place where the shift occurs. It makes no sense to ask which is the last red tube, not because there is such a tube but we cannot in principle know which it is (as in Timothy Williamson's 1994 account of these matters), but because there is no such thing. The question presupposes a determinate answer, but none exists.

How, though, is such indeterminacy possible? We can, perhaps, dispel any remaining mystery by recalling the semantic freedom that is the concomitant of Peirce's account of vagueness. Suppose one is called upon to answer the successive questions "Is tube a_1 red?," "Is tube a_2 red?," etc. What C records is the unsurprising fact that one will have to stop answering "yes" somewhere if one is to avoid giving an affirmative answer to the question

"Is tube a_{100} red?" Within limits, though, one may choose where to stop answering "yes." One has to stop somewhere, but there is no place where one has to stop. The verifying disjunct of C , then, is indeterminate because it is, within limits, *arbitrary*. When introduced to the Sorites, many people's first reaction is to say that there must be a switch in color somewhere, but there is no fact of the matter where, and the place we choose to make the switch does not mark any boundary of the concept *red*. Our analysis shows that this naïve reaction to the paradox is essentially right. Or better, the Peircean semantics provides a theoretical context within which this naïve thought can alleviate the sense of perplexity that the truth of C —the ineluctable conclusion of our Sorites argument—presents.

Where, finally, does this leave the Revised Argument for Bivalence, when this is applied to a vague statement? Unlike the supervaluationist, we can allow that a vague statement expresses just one thought: the statement A_{50} (e.g.) says that the tube a_{50} is red—no more, no less. We can also accept the relevant instance of excluded middle: we can say that the tube a_{50} is either red or not red. This too is attractive: a_{50} is problematic because it is indeterminate whether it is red; there is no impulse to deny that it is one or the other. All the same, the statement A_{50} is not bivalent. The Revised Argument to show that it is bivalent falls at the final hurdle, *viz.* in the application of \vee -elimination to reach line (16). In quantum logic, we cannot apply \vee -elimination in the presence of a non-logical premise; yet the assumption that the relevant statement satisfies (D), made at line (3), is such a premise. (The analogous argument to show that a true disjunction contains a true disjunct breaks down at the corresponding point.) In both set theory and vague discourse we find statements that are neither true nor false, but only vague statements put pressure on classical logic itself.²⁰

REFERENCES

- Andjelković, M. and T. Williamson (2000). "Truth, Falsity and Borderline Cases," *Philosophical Topics* 28: 211–44.
- Anscombe, G. E. M. (1956). "Aristotle and the Sea Battle," *Mind* 65: 1–15.
- Auxier, R. E. and L. E. Hahn (eds.) (2007). *The Library of Living Philosophers: Michael Dummett*. Chicago: Open Court.
- Baldwin, J. M. (editor-in-chief) (1901–5). *Dictionary of Philosophy and Psychology*. New York: Macmillan.
- Benacerraf, P. (1965). "What Numbers Could Not Be," *The Philosophical Review* 74: 495–512.
- Benacerraf, P. and H. Putnam (1983). "Introduction," In Benacerraf and Putnam (eds.) (1983), 1–37.
- Benacerraf, P. and H. Putnam (eds.) (1983). *Philosophy of Mathematics: Selected Readings*, 2nd edn. Cambridge: Cambridge University Press.
- Brouwer, L. E. J. (1927). "Intuitionistische Betrachtungen über den Formalismus." *Koninklijke Akademie van Wetenschappen te Amsterdam, Proceedings of the Section of Sciences* 31: 374–9.

²⁰ I am grateful to the editor of this volume for his comments on a draft.

Note added (July 2015). I completed this essay in 2012 and although I still think it publishable, my view of some of the matters it treats has evolved since I finished it. (1) I now prefer to justify the use of classical logic when reasoning about absolutely all sets by way of a negative translation into an intuitionistic language rather than via the supervaluational semantics proposed in sections 15.5 and 15.6. This alternative approach confirms the main thesis of those sections—that there is no general reason to expect the statements of set theory to be bivalent. (2) Contrary to sections 15.7–15.9, I now think that the Sorites paradox, and Wright's paradox of sharp boundaries, can be resolved without deviating from classical logic. For a defence of these claims, see Rumfitt (2015: chs. 9 and 8).

- Page references are to the English translation by S. Bauer-Mengelberg, entitled "Intuitionistic Reflections on Formalism," in J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press (1967), 490–2.
- Cohen, P. J. (1966). *Set Theory and the Continuum Hypothesis*. New York: W. A. Benjamin.
- Dummett, M. A. E. (1959). "Truth." *Proceedings of the Aristotelian Society* 59: 141–62. Page references are to the reprint in Dummett (1978), 1–19.
- Dummett, M. A. E. (1978). *Truth and Other Enigmas*. London: Duckworth.
- Dummett, M. A. E. (2000). *Elements of Intuitionism*, 2nd edn. Oxford: Oxford University Press.
- Dummett, M. A. E. (2007). "Reply to Crispin Wright," in Auxier and Hahn (eds.), 445–54.
- Ewald, W. B. (ed.) (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, vol. 2. Oxford: Oxford University Press.
- Feferman, S. (2009). "Conceptions of the Continuum," *Intellectica* 51: 169–89.
- Feferman, S. (2011). "Is the Continuum Hypothesis a Definite Mathematical Problem?" Draft of a paper read to the Harvard Philosophy Department in November 2011, and available as item no. 95 in the 'Papers' section of the late Prof. Feferman's homepage, which is still maintained on the Stanford Mathematics Department's website.
- Field, H. H. (1998). "Which Undecidable Mathematical Sentences Have Determinate Truth Values?" in H. Garth Dales and G. Oliveri (eds.), *Truth in Mathematics*. Oxford: Oxford University Press, 291–310.
- Fine, K. (1975). "Vagueness, Truth, and Logic," *Synthese* 30: 265–300. Reprinted in Keefe and Smith (eds.) (1996), 119–50.
- Glanzberg, M. (2001). "The Liar in Context," *Philosophical Studies* 103: 217–51.
- Gödel, K. F. (1940). *The Consistency of the Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*. Princeton: Princeton University Press.
- Gödel, K. F. (1964). "What is Cantor's Continuum Problem? (Revised and Expanded Version)," in Benacerraf and Putnam (eds.) (1983), 470–85.
- Goldblatt, R. I. (1974). "Semantic Analysis of Orthologic," *Journal of Philosophical Logic* 3: 19–35.
- Goldblatt, R. I. (1984). *Topoi: The Categorical Analysis of Logic*. Amsterdam: Elsevier.
- Heyting, A. (1934). *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie*. Berlin: Springer.
- Hintikka, K. J. J. (1964). "The Once and Future Sea Fight: Aristotle's Discussion of Future Contingents in *De interpretatione* IX," *The Philosophical Review* 73: 461–92.
- Hintikka, K. J. J. (ed.) (1969). *The Philosophy of Mathematics*. Oxford: Oxford University Press.
- Isaacson, D. R. (2011). "The Reality of Mathematics and the Case of Set Theory," in Z. Novák and A. Simonyi (eds.), *Truth, Reference, and Realism*. Budapest: Central European University Press, 1–75.
- Keefe R. and P. Smith (eds.) (1996). *Vagueness: A Reader*. Cambridge, MA: MIT Press.
- Kneale, W. C. (1972). "Propositions and Truth in Natural Languages," *Mind* 81: 225–43.
- Kneale, W. C. and M. Kneale (1962). *The Development of Logic*. Oxford: Oxford University Press.
- Koellner, P. (2010). "On the Question of Absolute Undecidability," in S. Feferman et al. (eds.), *Kurt Gödel: Essays for his Centennial*. Cambridge: Cambridge University Press, 189–225.
- Kreisel, G. (1967). "Informal Rigour and Completeness Proofs," in I. Lakatos (ed.), *Problems in the Philosophy of Mathematics*. Amsterdam: North-Holland, 138–71. Page references are to the partial reprint in Hintikka (ed.) (1969), 78–94.
- Lewis, D. K. (1988). "Vague Identity: Evans Misunderstood," *Analysis* 48: 128–30.
- Łukasiewicz, J. (1922). "O determinizmie," in J. Łukasiewicz, *Z zagadnień logiki i filozofii* ed. J. Śłupecki (1961). Warsaw: PWN, trans. Z. Jordan as "On Determinism," in Storrs McCall (ed.), *Polish Logic 1920–1939* (1967). Oxford: Clarendon Press, 19–39.
- McGee, V. (1997). "How We Learn Mathematical Language," *The Philosophical Review* 106: 35–68.

- Mackie, J. L. (1973). *Truth, Probability, and Paradox: Studies in Philosophical Logic*. Oxford: Oxford University Press.
- Menzel, C. (1986). "On the Iterative Explanation of the Paradoxes," *Philosophical Studies* 49: 37–61.
- Menzel, C. (2014). "Wide Sets, ZFCU, and the Iterative Conception," *Journal of Philosophy* 111: 57–83.
- Myhill, J. (1952). "A Derivation of Number Theory from Ancestral Theory," *Journal of Symbolic Logic* 17: 192–7.
- Parsons, C. D. (1974). "The Liar Paradox," *Journal of Philosophical Logic* 3: 381–412.
- Prior, A. N. (1971). *Objects of Thought*. Oxford: Oxford University Press.
- Rumfitt, I. (2011). "Ramsey on Truth and Meaning," in B. Morison and K. Ierodiakonou (eds.), *Episteme, etc: Essays in Honour of Jonathan Barnes*. Oxford: Oxford University Press, 213–45.
- Rumfitt, I. (2014). "Truth and Meaning," *Proceedings of the Aristotelian Society Supplementary Volumes* 88: 21–55.
- Rumfitt, I. (2015). *The Boundary Stones of Thought: An Essay in the Philosophy of Logic*. Oxford: Oxford University Press.
- Rumfitt, I. (2018). "Vagueness and Intuitionistic Logic," in A. Miller (ed.), *Logic, Language, and Mathematics: A Festschrift for Crispin Wright*. Oxford: Oxford University Press.
- Sainsbury, R. M. (1990). *Concepts without Boundaries*. London: King's College. Page references are to the reprint in Keefe and Smith (eds.) (1996), 251–64.
- Smiley, T. J. (1993). "Can Contradictions Be True?" *Proceedings of the Aristotelian Society Supplementary Volumes* 67: 17–33.
- Tarski, A. (1935). "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica* 1: 261–405, trans. J. H. Woodger as "The Concept of Truth in Formalized Languages," in A. Tarski, *Logic, Semantics, Metamathematics*, ed. J. H. Woodger and J. Corcoran (1983). Indianapolis: Hackett, 152–278.
- Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341–76.
- Uzquiano, G. (1999). "Models of Zermelo Set Theory," *The Bulletin of Symbolic Logic* 5: 289–302.
- Uzquiano, G. (2002). "Categoricity Theorems and Conceptions of Set," *Journal of Philosophical Logic* 31: 181–96.
- Varzi, A. C. (2007). "Supervaluationism and its Logics," *Mind* 116: 633–76.
- Walmsley, J. O. (2002). "Categoricity and Indefinite Extensibility," *Proceedings of the Aristotelian Society* 102: 217–35.
- Weston, T. (1976). "Kreisel, the Continuum Hypothesis, and Second-Order Set Theory," *Journal of Philosophical Logic* 5: 281–98.
- Williamson, T. (1994). *Vagueness*. London: Routledge.
- Williamson, T. (1998). "Indefinite Extensibility," *Grazer Philosophische Studien* 55: 1–24.
- Wright, C. J. G. (2007). "Wang's Paradox," in Auxier and Hahn (eds.), 415–44.
- Wright, G. H. von. (1984). "Determinism and Future Truth," in G. H. von Wright, *Truth, Knowledge, and Modality: Philosophical Papers Volume III*. Oxford: Blackwell, 1–13.
- Zermelo, E. F. F. (1908). "Untersuchungen über die Grundlagen der Mengenlehre I," *Mathematische Annalen* 65: 261–81, trans. S. Bauer-Mengelberg in van Heijenoort (1967), 199–215.
- Zermelo, E. F. F. (1930). "Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre," *Fundamenta Mathematicae* 16: 29–47. Page references are to the translation by M. Hallett, "On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory," in W. B. Ewald (ed.) (1996), 1219–33.

CHAPTER 16

TRUTH, OBJECTIVITY, AND REALISM

SANFORD SHIEH

16.1 INTRODUCTION: METAPHYSICS AND TRUTH

PHILOSOPHY has perennially been exercised by metaphysical questions about the existence and nature of various types of entity. Here is a sampling of familiar ones:

- Do moral and aesthetic values exist at all, and if so, do they exist independently of thoughts and practices of human beings?
- Do numbers and geometrical figures exist independently of particular concrete physical objects or of our mathematical thought?
- Do medium-sized specimens of dry goods exist independently of being perceived?
- Do (other) minds exist, or are mental states nothing more than dispositions to behave in certain ways, or states of the central nervous system?
- Is time real; do the past and the future exist now?

These questions have of course engendered detailed and intricate answers. But it is not a gross distortion to classify the answers into two groups of opposed metaphysical positions. On the one side are various types of realism. These positions affirm that the entities in question exist, i.e. are a part of reality, or, more strongly, that they possess one of two types of independent existence. One type is exemplified by one answer to the question about moral values: they exist, and would exist even if we human beings did not take anything to be morally good or right; moreover, whatever is good or right would be so whether or not we took them to be so. This is to take moral values to be *objective* aspects of reality. Another type of independence is an answer to the question about minds: mental states exist, and are genuinely distinct from, not reducible to,

behavioral dispositions. This is to take the mental to be a (relatively) *fundamental* aspect of reality.

On the other side are positions which deny that the entities in question exist independently, i.e. are an objective or a fundamental part of reality, or, more strongly, that there are such entities at all. For example, one answer to the question about aesthetic values is a style of relativism which claims that an object is beautiful or ugly only in virtue of being taken to be so by an individual, or an institution, or a culture. Aesthetic properties, on such a view, exist, but only by dint of the attitudes or perspectives of an individual or group, and so are not *objective* features of reality. Another example is an answer to the question about minds: mental states exist, but are at bottom nothing more than states of the central nervous system; they are not *fundamental* features of reality but *reducible* to states of certain physical objects. Finally, J. L. Mackie (1977) advances an error theory about morality claiming that there are no moral values at all, that morality is no feature of reality.

Note that the two types of independent existence are, *prima facie*, not equivalent. It seems consistent to hold that mental states are states of the central nervous system, and so, like states of any other physical object, exist objectively, independently of being taken to exist or to be instantiated. On this combination of views the mind is an objective but not fundamental aspect of reality.

This classification is, of course, very rough. Not all traditional metaphysical positions can be fitted comfortably into it. In particular, Berkeley's idealism about physical objects of our ordinary experience can be interpreted in two ways in this scheme. One can take the view to be that such objects exist and have the properties they have only in virtue of being perceived and taken to be so by human or divine beings, and so are not objective aspects of the world. One can also take the view to be that such objects are at bottom nothing more than human and divine ideas, and so are not irreducible, fundamental parts of reality.

How are these metaphysical positions related to the notions of truth and falsity? One view is this. Consider the main questions of the classic debates over truth at the turn of the twentieth century: what entities are capable of being true, and what does the truth of these entities consist in? The answers proposed to the second question are generally analyses of the concept of truth in independent and prior terms. Correspondence theories, for a central instance, held that those things apt to be true or false, such as thoughts and statements, are true or false in virtue of standing in a relation of correspondence to some aspect of reality. What exactly is the relation of correspondence, and what exactly are the aspects of reality to which thoughts and statements correspond, are intricate questions.¹ But however these are answered, it seems that a statement for example can be true only if those aspects of reality to which it corresponds exist, and the truth of that statement is objective or not depending on whether those aspects exist objectively. On

¹ See especially Walker (ch. 8), David (ch. 9), Candlish and Damnjanovic (ch.10), and Misak (ch. 11), all in this volume.

this picture, the metaphysics of truth is determined by the metaphysics of that in terms of which truth is analyzed.

Falsity in correspondence theories is trickier, but doesn't seem to complicate this basic picture. Begin from the intuitive idea that statements purport to be about certain entities. "Netrebko sings" for instance, purports to be about a certain soprano and her primary professional activity. If in reality Netrebko does not sing, the statement is of course false. One might say that this falsity is due to a fact's failing to obtain in reality, although it might have obtained; that fact is, of course, what our statement purports to represent. And, the objectivity of this statement's falsity would be measured by the independence or otherwise of the failure of that possible fact to obtain. But what if in reality, or objective reality, there is no Netrebko? In this case of course no fact in reality obtains to make the statement true, but the absence of Netrebko means that there couldn't have been a fact whose obtaining would have made the statement true. Since the statement purports to represent a fact in which Netrebko figures, in such a case one might say that we are under a misapprehension about our discourse: the ontological presuppositions required for our statement to be a representation of the world, correct or incorrect, are not (objectively) satisfied. There are then three familiar options. One might take this statement to be (objectively) false. One might take it to have no (objective) truth-value. Or, one might take it not to be (objectively) a genuine statement at all, not capable of being true or false, but merely a sentence or utterance. Let's use the term "truth-status" of a statement to refer to its truth-value or its failure to have one. Then we can say that, whichever option one takes, the truth-status of a statement—true, false, or neither—and the capability of an utterance to have a truth-status, are aspects of reality determined by the metaphysics of facts and of entities. Hence the concepts of truth and falsity have no special metaphysical significance.

This chapter is concerned with a contrasting view, initially advanced by Michael Dummett and subsequently developed by Crispin Wright. They propose taking the concept of truth as fundamental, and explicating metaphysical debates about realisms in terms of truth. The heart of their proposal is to replace traditional metaphysical questions about existence with questions about the truth conditions of corresponding classes of statements: mathematical, moral, and aesthetic statements; statements apparently about physical entities and mental states; and statements in the past and future tense. (I will occasionally follow Wright's practice of calling such a class of statements a "discourse.") Call Dummett's and Wright's proposal the semantic (as opposed to ontological) approach to metaphysics.

Dummett originally goes on from this basic move to characterizing realism with respect to an aspect of the world as the thesis that each of a class of statements "possess[es] an objective truth-value, independently of our means of knowing it" (1963: 146). Thus, while ontological realism about mathematical objects, sometimes called platonism, claims that numbers and figures are abstract non-spatio-temporal entities that exist independently of our mathematical thought, semantic realism about mathematics holds that arithmetical or geometrical statements are true or false independently of our capacities to recognize their truth-value. Similarly, while ontological nominalism

about mathematics holds that mathematical objects are creations of our mathematical practice, semantic anti-realism about mathematics holds that arithmetical and geometrical statements can have truth-values only if we are (at least in principle) capable of recognizing that they have those truth-values.

The semantic approach should be distinguished from two much-discussed recent styles of opposition to realism that also focus on classes of statement. One of these, error-theories, is based on the sort of failures of ontological presuppositions of statements mentioned above. Two prominent contemporary instances are Mackie's view that since there are no moral properties, all moral statements are false, and Hartry Field's (1980) view that since there are no mathematical objects, all mathematical statements are false. The other, expressivist theories, also begin from failures of ontological presupposition. But, instead of concluding that our uses of moral or mathematical vocabulary are uniformly statements of falsehoods, expressivisms conclude that these uses are not the making of statements but the expressions of attitudes. The classic instance of expressivism is emotivism about ethics, on which moral statements are expressions of approval or disapproval. What distinguishes these anti-realisms from the semantic approach is that, in both cases, ontological views about the purported subject-matter of the discourses in question determine semantic properties of the discourses, rather than the other way around.

It is also important to distinguish two aspects in Dummett's original characterization of realism. First, realism consists of the *objectivity of truth-value* of a discourse. Second, the objectivity of truth-value is explained in terms of the *independence* of the properties of truth and falsity from our *capacities to recognize* that statements of the discourse have those properties. This feature of truth is often called "recognition transcendence." What is fundamental to Dummett's and Wright's proposals is taking realism to consist in the objectivity of truth.² But it is not mandatory in the semantic approach for the objectivity of truth to be taken as the recognition transcendence of truth. This point is fairly clear in the case of Wright, who was moved by dissatisfaction with Dummett's original recognition transcendence criterion to advance three other objectivity-constituting characteristics of truth. It is less commonly appreciated that Dummett, in (1981a; 1982), also came to hold that his original criterion of objectivity should be supplemented by two other criteria.

From this last point it should be clear that the notion of objectivity at work in the semantic approach is not the same as that involved in the ontological approach. Objectivity of truth-value is the objectivity of linguistic representation, and, on Dummett's and Wright's views, can be constituted in a variety of ways. Objectivity of existence, by contrast, is always existence independent of representations of existence.

² Dummett's and Wright's proposals for non-metaphysical construals of realism are closely related to Hilary Putnam's internal realism. But considerations of space, together with the fact that Putnam focuses on reference rather than truth, led to me leave out a discussion of his views. For the relation between Putnam's and Dummett's views see Putnam (1983: Introduction) and Dummett (1991). For Wright on Putnam see Wright (1992: 38–47) and Wright (2003: ch. 11); see also Putnam's (2001) reply to Wright.

I have two main aims in this chapter. First, I want to give a sense of what is distinctive about the semantic approach, by providing an overview of Dummett's and Wright's arguments and conclusions. In sections 16.2 and 16.3 I discuss Dummett's views. I begin in section 16.2 with an account of the three Dummettian criteria for the objectivity of truth. It will emerge here that these criteria furnish a more subtle and flexible framework for characterizing realism than that which is usually attributed to Dummett in the secondary literature. In particular, we will see that for Dummett realism about a discourse comes in degrees, and that, perhaps rather surprisingly, neither Frege's view of singular terms nor neo-Fregean views of arithmetic count as fully realist. In section 16.3 I turn to the most notorious of Dummett's semantic metaphysical arguments, the so-called "manifestation" arguments against semantic realism in general. Here I distinguish two interpretations of the argument, one based on a "substantial" view of knowledge of meaning as a psychological state underlying linguistic competence, and the other based on a "fictionalist" view of knowledge of meaning as a theoretical description of practical linguistic abilities having no psychological reality. Many objections to the argument work only against the substantial interpretation, but the most significant objection is relevant to both. Drawing on Wright's response to this objection, I argue that it can be answered on the basis of some intuitions about the relation between knowledge and recognitional capacity, and so the soundness of the manifestation argument remains an open question. In sections 16.4 and 16.5 I discuss Wright's views; the treatment here is relatively briefer since Wright's semantic metaphysics has been the subject of much recent work.³ Wright's program rests on a critique of deflationary theories of truth that motivates his minimalist theory of truth. In section 16.4 I argue that, to make sense of Wright's critique, deflationism must be taken to include a rejection of any ontological metaphysical grounding of the concept of truth. Moreover, Wright's minimalism also incorporates this rejection, and so is, in a sense, a continuation of deflationism by other means. In section 16.5 I outline Wright's four semantic criteria of realism and examine their inter-relations as well as their connections to Dummett's criteria. I show that two of the criteria Wright proposes in addition to Dummett's recognition transcendence share the broad motivations, but not the details, of the two additional Dummettian criteria I uncover in section 16.2. Throughout these four sections I generally do not examine the steps of the arguments in close detail, nor do I attempt to assess their soundness.

My second main aim is to give an account of what might be called the meta-philosophical status of the semantic approach. That is, I'm interested in the question:

- What kind of philosophical investigation is undertaken through the semantic approach, and (how) does it differ from the ontological approach?

³ See, *inter alia*, Jackson (1994); Williamson (1994; 1996); Van Cleve (1996); Horgan (1996); Horwich (1996); Pettit (1996); Sainsbury (1996); Blackburn (1998); and Taschek and Shapiro (1996).

In section 16.6 I approach an answer to this question by discussing the motivation and adequacy of semantic metaphysics, i.e. I discuss the following questions:

- Why pursue the semantic rather than the ontological approach?
- Is the semantic approach adequate to metaphysics? Are there legitimate metaphysical issues that cannot be addressed semantically?
- Does the semantic approach lead to metaphysical conclusions that are intuitively or obviously incorrect?

I begin by outlining and responding to a well-known two-part critique of Dummett's views, originally advanced by Michael Devitt (1983; 1991: ch. 14) and further developed by Alexander Miller (2005: section 6; 2006). They take Dummett to justify the semantic approach on the basis of the claim that traditional ontological theses are merely metaphorical, so that the only version of realism with literal content is semantic, and argue that Dummett fails to justify this claim. Against this I outline a better construal of the semantic approach, as a proposal recommended on the basis of certain advantages, rather than as resting on a thesis about how metaphysics has to be conducted. Devitt also argues that the semantic approach is intuitively inadequate since sense-data or Berkeleian idealism could count as semantically realist. Against this I argue that Berkeleian idealism could count as ontologically realist, if realism consists in claims of objective existence in the sense outlined in this section. I conclude with a discussion of a general two-part worry about the semantic approach raised from the ontological perspective: the semantic approach rules out (1) the possibility of facts in reality that are not fulfilled truth conditions of our statements or thoughts, and (2) the possibility that large classes of what we take to be statements or thoughts fail to depict any facts in reality. I argue that the semantic approach need not rule out the first sort of possibility, but merely hold that those possible facts are not amenable to interesting metaphysical investigation. I also argue that Wright's responses to expressivisms and error-theories constitute an answer to the second sort of possibility: the intuitive examples on which these possibilities are based can all be reconstrued as cases of statements which fail to satisfy our standards for evaluating the correctness of assertions.

16.2 DUMMETT'S SEMANTIC CRITERIA FOR REALISM

Our main concern in this section is the relation between recognition transcendence and the other criteria for realism resulting from Dummett's revised view of realism. But I begin with a discussion of what seems to be another characteristic commitment of realism for Dummett. He continues the formulation of realism quoted above—statements “possess an objective truth-value, independently of our means of knowing it”—by

saying that these statements “are true or false in virtue of a reality existing independent of us” (Dummett 1963: 146). This suggests that realism about a discourse involves

Strict bivalence: Each statement of that discourse is either true or false.

What is the relation between recognition transcendence and strict bivalence? Are they independent features jointly characterizing realism? Or is one constitutive of realism, and the other merely a consequence of realism? A plausible view is that, if the statements in question include what Dummett calls “undecidable” or “not effectively decidable” ones, then strict bivalence suffices for recognition transcendence. In order to see why, we first have to explain the notions of decidable and undecidable statement.⁴ A statement is decidable if we know we can either recognize it to be true or recognize to be false. Clearly we would be in this position if we know of a decision procedure for that statement, which is guaranteed to enable us, in principle, to tell whether it’s true or false. A statement is undecidable if it’s not decidable, i.e. if we *don’t know that we can* recognize it to be true or to be false. Note that an undecidable statement is *not* a statement that we *cannot* recognize to be true or false. Decidability and undecidability are temporal epistemic notions: a sentence undecidable at some time could be decidable at a later time, if a verification happened to come along.

Suppose now that an undecidable statement *S* satisfies strict bivalence. Since *S* is either true or false, the condition for its truth either obtains or fails to. Since *S* is undecidable, we don’t know that we can recognize *S* as true or as false, i.e. we don’t know that we can recognize that the condition for *S* to be true obtains or that it fails to obtain. It follows that, for all we know, the truth condition of *S* can (fail to) obtain although we cannot recognize that it does. Note that here we are construing the possibility that truth transcend our capacity for recognition as epistemic possibility.

What about the converse? Is strict bivalence necessary for recognition transcendence? Dummett’s main reason against this is Frege’s rejection of strict bivalence for statements containing non-referring singular terms. On Frege’s view such a statement is neither true nor false, but its being in this state is independent of our recognitional capacities; it may be in principle impossible for us to recognize whether the sense of a singular term succeeds or fails to determine a referent. Frege’s view is that strict bivalence fails, but all statements have a determinate truth-status that is recognition transcendent. Recognition transcendence of truth-status does not seem a departure from realism; the truth state of a statement is still fixed by reality unconstrained by our cognitive powers. Hence Frege’s rejection of strict bivalence for truth-value gaps appears metaphysically shallow.

If this is right, then we might take recognition transcendence, applied to truth-status, to be the fundamental criterion for realism, and strict bivalence to be merely one path

⁴ For more extensive discussion of this account of undecidability in semantic anti-realism, and its differences from other conceptions of undecidability, see Shieh (1998a).

to realism, sufficient to secure realism but not necessary for it. This view seems also to cohere with what is usually taken to be Dummett's argument against classical logic. The idea is that classical logic is underwritten by bivalence, so rejecting bivalence deprives classical logic of one avenue of justification. But if bivalence implies realism, then arguments against realism would undermine bivalence.⁵

Dummett's later view of realism adds a further condition:

[To] have a realistic view, it is not enough to suppose that statements of the given class are determined, by the reality to which they relate, either as true or as false; one has also to have a certain conception of the manner in which they are so determined. This conception consists essentially in the classical two-valued semantics: and this, in turn, embodies an appeal to the notion of reference as an indispensable notion of the semantic theory. (Dummett 1982: 231)

Moreover, at this point Dummett explicitly disavows the account of the relation between realism and strict bivalence just canvassed. Strict bivalence, on the later view, is necessary but not sufficient for semantic realism. Rather, strict bivalence and classical semantics are individually necessary and jointly sufficient for realism.

What is classical semantics? For our purposes we can characterize it as a theory of truth-value determination for a(n extensional first-order) language with two features. First, the truth-values of all logically complex sentences, including quantified sentences, are fixed by the truth-values of the atomic sentences of either that language or an expansion of it containing singular terms referring to all objects in the (first-order) domain of quantification. Second, the truth-value of an atomic sentence is determined by whether an appropriate ordered sequence of the referents of the singular terms occurring in that sentence is a member of the extension of the predicate occurring in that sentence. What's crucial is that according to classical semantics the referents of singular terms play an ineliminable role in the truth conditions of all sentences. This last point might be taken to show that this semantics is Fregean rather than Tarskian, because in the standard Tarskian formulation of first-order semantics it is satisfaction of open sentences, rather than reference of constants, that play the crucial role; there may be no constants in languages to which Tarskian semantics applies.⁶ But an open sentence is satisfied by a sequence of objects just in case the closed sentence resulting from replacing the free variables in the open sentence with constants referring to the objects of the sequence assigned to the variables replaced is true. So it's not clear that Tarskian first-order semantics is, from the perspective of the philosophical issues we're concerned with, anything more than a notational variant of Fregean classical semantics.⁷

⁵ See Shieh (2012) for a different view of the appropriate anti-realist basis for a criticism of classical logic.

⁶ Of course in Tarski's work in model theory the cardinality of the set of constants of the languages treated is crucial. See e.g. Tarski (1954–5).

⁷ Dummett makes this argument in (1981b: 16–17).

Why does strict bivalence by itself fail to guarantee realism? The reason is *not* that Dummett has given up the argument from strict bivalence for undecidables to recognition transcendence of their truth-values. The reason is, rather, that a semantic theory incorporating recognition transcendent strict bivalence may nevertheless be non-classical. An example is the neo-Fregean view of abstract objects advanced by Wright and Bob Hale.⁸ They hold that all that's required, in order for a class of singular terms to refer to objects, is that these terms occur in sound inferences conforming to certain (first- and higher-order) logical laws, and that statements in which these terms occur, most importantly identity statements, have determinate truth conditions. A central case, discussed by Frege, consists of terms for directions. The truth conditions of identities between direction terms are specified by statements ascribing parallelism to lines, and those of atomic predications of direction terms are specified by predications of terms for lines. Dummett does not dispute that these neo-Fregean logico-syntactic conditions on expressions suffice for these expression to be singular terms that refer to objects. But he points out that the reference of these singular terms, explained in this way, plays no role in the account of how sentences containing these terms are determined as true or false. In the case of direction terms, for example,

the canonical means by which we establish [an] atomic sentence [containing a direction term] as true or as false is not by identifying some direction as being that to which the term refers, and then determining that the predicate is true of it; it is by first translating the sentence into a statement about lines, and then determining, by whatever are the appropriate means, the truth-value of the resulting statement. (Dummett 1982: 241)

Why, though, does this show that the neo-Fregean view is not realism?

We can begin to see why by considering the contrast between the neo-Fregean view and reductionism about directions:

If . . . a term for a direction [is] replaced by a term for . . . a maximal class of parallel lines, then . . . terms for directions [are] construed as referring to classes of parallel lines. In this . . . case, therefore, we should have a reductionist thesis that in no way impugned a realistic interpretation of statements about directions. (Dummett 1982: 246)

This contrast is based on a criterion for realism distinct from recognition transcendence. The idea at work here is that to conceive of a class of entities as fully a part of reality is to take facts about these entities as that which make our statements—our attempts to represent the part of reality containing these entities—true or false. But to think of facts about some entities as making our representations true or false requires that there be elements in our representations that correspond to, i.e. that refer to, these entities. In

⁸ See Wright (1983) and Hale and Wright (2001).

the initial neo-Fregean construal of direction statements, direction terms are supposed to have reference. But it turns out that it's not facts about these referents of direction terms that fix the truth-value of direction statements. Rather, it's facts about lines that fix these truth-values. So, even if we take it that there are directions, i.e. that we refer to directions, we can't take our direction statements to be, at bottom, depictions of *these* entities. So, on this construal, we are conceiving of directions not as what's ultimately there, in reality, but as objects that are in some way dependent on what's ultimately real, namely, lines. All this changes on the reductionist semantics for direction terms. For here we take it that direction terms refer to classes of lines; that is, directions *are* classes of lines. Moreover, on this construal facts about classes of lines make our representations of directions true or false, in that since directions are classes of lines, facts about directions are the truth-makers of our direction statements.

Neo-Fregean semantics for a class of statements represent a departure from classical semantics that retreats from the view that those statements "relate to some reality that renders those statements true or false." The stress here falls on "renders the statements true or false." Neo-Fregeanism about directions takes direction statements to relate to some reality, namely the realm of directions to which direction terms refer. But, on that view, the reality containing directions is not that which renders direction statements true or false.

The criterion for realism we have just discussed allows us also to understand why, on Dummett's later conception, Fregean truth-value gaps represent a rejection of realism. On the later conception, an anti-realism is always "correlative to a corresponding species of realism" (Dummett 1982: 265), and each non-classical semantic theory is opposed to some "possible form of realism . . . embodied in a truth conditional meaning-theory based on a two-valued semantics" (1982: 266). In particular, Frege's truth-value gap view is anti-realist in relation to "realism of a Meinongian kind, which would take all proper names as referring to objects, whether existent or non-existent, and would construe statements of the form '*a* is *F*' as meaning '*a* exists and is *F*'" (1982: 269). Note that Russell's theory of descriptions also represents a kind of anti-realism in relation to Meinongian realism. This is because Russell, in contrast to Meinong, takes most ordinary proper names not to be singular terms at all, so that they don't have referents which play a role in how sentences in which they occur are determined as true or false (*ibid.*).

The important point is that both Frege and Russell, in contrast to Meinong, take some expressions, which appear to be singular terms, to contribute something other than their referents to the truth conditions of sentences in which they occur. So, just as in the case of neo-Fregeanism about directions, what render these sentences true or false are not facts concerning a realm of entities picked out by the apparent singular terms in question. On Frege's view, the terms pick out nothing. On Russell's view, it's facts about the properties expressed by the description, rather than any entity correlated with the description, that fixes the truth-values of the sentences. On Meinong's view, in contrast, each singular term picks out something, and facts about those entities make the sentences in question true or false.

The upshot of this discussion is that on Dummett's considered view semantic realism is more than recognition transcendence of truth, of truth-statuses, or truth conditions. A second criterion that we've just unearthed is that truth conditions involve the referents of (apparent) singular terms. What I turn now to show is that there is yet a third criterion, which appears in the initial characterization in the phrase "determinately."

In order to see this, let's look at the simplest of a family of semantic positions Dummett calls neutralism about the future (1982: 257). This is a familiar supervaluational semantics in which the (present) truth-value of a future-tense statement is determined with respect to a set of alternative future states of the world. Each future tense statement is conceived of as the result of applying a future tense operator—say, "*F*"—to an untensed statement *p*. With respect to each possible future state, *p* is determined as either true or false. "*Fp*" counts as presently true just in case *p* is true with respect to all possible future states of the world, presently false just in case *p* is false with respect to all future states. Strict bivalence of course can fail, but the law of excluded middle and weak bivalence both hold.⁹

So, now, don't we have a position exactly analogous to Fregean truth-value gaps? Here again we have a failure of strict bivalence with a holding of weak bivalence. Here, also, we seem to have a semantics in which all future tense statements have completely determinate truth-statuses. Dummett claims that this type of neutralism is anti-realist. So our question is, can we apply the explanation of the anti-realism of Fregean truth-value gaps to neutralism? It seems not. The reason is that it seems perfectly consistent to be a Meinongian about apparent singular terms and yet a neutralist about the future. All states of the world, including the single past and present state and all possible future states, contain non-existent entities to which "empty" singular terms refer. Untensed statements are rendered true or false, with respect to each future state, by facts involving the referents of all apparent singular terms occurring in them. What, then, makes neutralism anti-realist?

Dummett's answer is not very helpful: "We shall certainly regard the neutralist as an anti-realist, since the conception under which every statement about the future already has a determinate truth-value exerts a strong attraction" (1982: 267). Whether this conception is indeed attractive may well be questioned; what's important, though, is the nature of the conception. The realism from which neutralism retreats is the view that reality contains but a single future state of the world, with respect to which the truth-values of future tense statements are fixed. What underlies this conception of realism is the Tractarian idea that reality is the totality of facts, i.e. states of affairs that *actually* obtain. Neutralism represents a rejection of this criterion of realism because it rejects the view that there is, presently, a single totality of future facts that actually obtain. This is what it means to claim that there is no determinate future reality to render future tense statements true or false. Put semantically, on this view future tense statements do not depict a single future reality.

⁹ The canonical statement of this semantics is, of course, Thomason (1970).

Note that there would be no indeterminacy if distinct alternative possible future states of the world are conceived as parts of a more inclusive totality of facts. Is there such a conception of the future? There certainly is such a conception of modality: David Lewis's modal realism.

The determinacy criterion of semantic realism allows us to make some sense of some aspects of Dummett's view of realism and anti-realism with respect to the phenomenon of vagueness.

Vagueness has seemed to some to pose a decisive problem for the view that any rejection of strict bivalence is a departure from (full) semantic realism. If one finds epistemicism—the view that ascriptions of a vague predicate are invariably true or false, even if made to an element of the penumbra of that predicate—hard to accept, then one is committed to some of these ascriptions being truth-value gaps. But it seems consistent with this commitment to hold that the facts (whatever they are) depicted by vague statements obtain independently of us, of our cognitive powers and linguistic practices.¹⁰

The answer to this objection begins from the observation that recognition transcendence is not the only criterion of realism. Non-epistemicist conceptions of vagueness are not fully realist, not by being recognition immanent, but by departing from the determinacy criterion of realism.

Full realism about vague predicates is epistemicism. Considered semantically, epistemicist views amount to the claim that the condition in which a vague statement is true is exactly that in which the object picked out by the singular term(s) is in the extension of the vague predicate, and *there is no vagueness* in whether any object is in the extension of that predicate. So, not only do vague statements conform to strict bivalence, but they have a two-valued classical semantics. But note that this realism about vague statements is tantamount to a denial of any vagueness *in the states of affairs* making these statements true or false. Since determinacy of truth conditions is necessary for realism, realism with respect to vague statements is a variety of eliminativism about vagueness in reality.

A rejection of epistemicism involves some sort of modification of classical semantics for vague statements. One standard type of modification is a supervaluationism formally identical to neutralism: instead of various possible alternative futures states of the world, we have here a range of possible "precisifications" of a vague predicate. An atomic vague statement " Fx " is true if the referent of " x " is in the extension of every precisification of " F ," false if it is in the anti-extension of every precisification of " F ." Clearly if the referent of " x " is in the penumbra of " F ," strict bivalence fails for " Fx ."

The question now is, where is the indeterminacy of truth conditions in supervaluationism about vague statements? If we replaced the vague predicate with any of its precisifications, then we would obtain a statement made true or false by the obtaining or otherwise of the state of affairs it depicts. But the truth-value of the original vague statement is not fixed by the obtaining of any of *these* states of affairs. None

¹⁰ See Rosen (1995: 605).

of these states of affairs, then, is what the vague statement determinately represents. The indeterminacy here lies in what a vague statement represents: what such a statement represents is compatible with, but distinct from, what any of its precisifications represent. Thus, note that, compared to epistemicism, supervaluationism allows more reality to vagueness, but what is vague or indefinite lies in what our statements represent or say, not in states of affairs that obtain or fail to.

Note also a contrast with neutralism. In that case a single state of affairs would render a future tense statement not containing any vague expressions true or false, but from the vantage point of the present the future does not contain just a single state of affairs. Neutralism and supervaluationism about vagueness are both anti-realist because of indeterminacy in how truth-values are fixed. But the sources of indeterminacy differ. In neutralism it comes from the truth-maker of statements; in supervaluationism it comes from how statements represent.

There is an even more anti-realist view of vague statements according to which they depict indefinite states of affairs. But this view is also represents a fully realist attitude toward vagueness as aspects of reality. Much of what is puzzling in Dummett's claims about the phenomenon of vagueness lapses once one realizes that the strength of realism about vague *statements* varies inversely with the strength of realism about vagueness *in reality*, i.e. in how vague statements are rendered true or false.

To sum up, then, there are three criteria for realism about a region of discourse in Dummett's considered view:

- The conditions under which the statements are true involve the referents of the singular terms occurring in these statements.
- A condition in which the statement is true is a single definite condition, rather than a range of conditions each of which is compatible with what the statement represents.
- Statements have recognition transcendent truth conditions.

Corresponding to these ingredients of realism, there are (at least) three grades of anti-realism, three grades of retreat from a full realist semantics meeting all three criteria:

- An anti-realism in which the single determinate condition in which a statement is true involve entities other than the referents of the singular terms occurring in these statements. This anti-realism, exemplified by Frege's and Russell's accounts of statements containing (apparently) empty singular terms, is the mildest grade.
- An anti-realism which takes a statement to be true when a range of conditions, each of which is compatible with what the statement represents, exhibits some feature. This is exemplified by neutralism about the future, supervaluationism about vagueness, and one account of set theoretic statements. This is the second grade of anti-realism.
- The strongest grade of anti-realism takes statements to have truth conditions that are epistemically constrained.

I have just argued that, for Dummett's semantic approach, the realism/anti-realism distinction does *not* line up with the recognition transcendence/recognition immanence distinction. But since Wright, along with most of the literature, identifies them, to avoid repeated qualifications I will do likewise, for the most part, in the remainder of this chapter.

16.3 THE MANIFESTATION ARGUMENT FOR SEMANTIC ANTI-REALISM

In this section I discuss one of the most distinctive features of Dummett's semantic metaphysics: his "manifestation" argument that considerations about linguistic understanding lead to anti-realism about all discourses. This argument has been the subject of much critical scrutiny, and something like the following picture of how it runs is familiar. It begins from the Wittgensteinian view that meaning is use, the use of linguistic expressions in public circumstances. Wright characterizes this view as a "harmless platitude," on which linguistic understanding is a "complex of abilities, rather than . . . some kind of interior state" (1993: 247, 16). Dummett argues for it, on the basis that otherwise communication by language is impossible. The conclusion is a "manifestation principle": the state of understanding a sentence must be "exhaustively determined" by a speaker's observable uses of that sentence (1973: 216). The argument then continues by using this principle to make trouble for semantic realism. If the meaning of a sentence *S* is its truth conditions, and if these conditions can be recognized as obtaining when they do obtain, a speaker can manifest her knowledge of *S*'s meaning by asserting it when she recognizes it to be true and rejecting it when she recognizes it to be false. But if *S*'s truth conditions are undetectable, she cannot display her knowledge of those conditions by exercising such a recognitional capacity. There is, moreover, no other way for her association of one set of truth conditions with *S* as opposed another to appear in her use of language. Hence if a speaker takes *S* to be true in conditions that transcend recognition, it would be impossible for her to manifest that understanding of *S*. By the manifestation principle, it follows that the meaning of *S* cannot be such truth conditions.

In the mass of critical discussion devoted to this argument, one can discern two tendencies, corresponding to contrasting conceptions of knowledge of meaning, of use, and of the relation of manifestation. On the one hand, there are those who take what I'll call a *substantial* view of knowledge of meaning. On this view, a speaker's knowledge of the meanings of expressions is a psychological state that underlies his linguistic competence, in the sense that his uses of those expressions is guided by his knowledge of what they mean.¹¹ Observable uses provide evidence of psychological state, but do not constitute linguistic competence; it is the state of understanding that explains the speaker's abilities to use language, not vice versa. Manifestation, on this conception, is

¹¹ See Strawson (1977: 16); Devitt (1983: 83–90).

an evidentiary relation: use manifests understanding in the sense of providing evidence for it. Something like this conception seems to be present in Dummett's argument that the possibility of communication demands the manifestation of meaning. Linguistic communication is "conveying" our beliefs, desires, and other mental states to one another. In order for *A* to communicate something successfully to *B*, *B* has to know the meanings that *A* associates with the words she utters (in speech or writing). But, in order for *B* to attain this epistemic achievement, *A* has to provide *B* with evidence, in the form of *A*'s "uses" of linguistic expressions, for what *A* means by those expressions. That is to say, meaning has to be manifestable. Moreover, Dummett assumes that there is no more to meaning than what can be communicated. It follows that meaning has to be "exhaustively" manifestable. That is, if it is not possible to establish, on the basis of a speaker's uses of words, that she associates some meanings with those words, then there simply aren't any meanings attached to her utterances of those words.

On the other hand, there are those who take what I'll call a *fictionalist* view of knowledge of meaning. On this view, what fundamentally constitutes linguistic competence is, as Wright puts it, a complex of abilities to use expressions, rather some "interior" psychological state. Knowledge of meaning has no psychological reality, or, at any rate, does not guide a speaker's uses of language, and so plays no role in explaining her linguistic abilities. Knowledge of meaning, on this view, is implicit knowledge. A clear illustration of this notion is Richard Kirkham's example of the ability to touch type (1989: 36). Most touch typists can't, without looking at the keyboard, make a picture of it showing where all the keys are, or even tell whether a picture they are given is accurate or not. So, plausibly, they do not know a set of propositions stating the location of the keys. Yet they move their fingers (mostly) *as if* they had that knowledge, *as if* they inferred from those propositions how to move their fingers. In this sense, they have implicit knowledge of those propositions. Analogously, a theoretical representation of a speaker's linguistic abilities in terms of a set of propositions constitutes implicit knowledge just in case she uses words *as if* she knew these propositions. On this view, the ascription of implicit knowledge is a theoretical postulation to explain observable phenomena. Manifestation remains an epistemic constraint, based on Occam's razor: use manifests meaning in the sense that one should not postulate more propositions known by speakers than is required to account for what they can be observed to do with language. As Dummett puts it, "[i]mplicit knowledge cannot . . . meaningfully be ascribed to someone unless . . . there is an *observable* difference between the behaviour or capacities of someone who is said to have that knowledge and someone who is said to lack it" (1973: 225).¹²

There are two main lines of criticism directed against the manifestation argument. One questions the substantial view of meaning, on the basis of two types of argument. Some hold that the manifestation argument confounds what should be kept apart, namely, the theory of the meanings of expressions and the theory of speakers'

¹² For a contrasting, non-epistemic conception of manifestation see Shieh (1998b). With little explanation to go on, it's tempting to understand Wright's view of the manifestation principle in terms of this conception of implicit knowledge of meaning. Specifically, isn't Wright's claim that understanding

knowledge of meaning.¹³ The first is descriptive semantics, an attempt to specify what in fact are the semantic values of expressions of a language. The latter is foundational semantics or meta-semantics, an attempt to specify the facts about speakers in virtue of which the language they speak has whatever semantics it in fact has. Knowledge of meaning is a state of speakers, and its theoretical role, accordingly, is to explain how it is that the language of these individuals has the semantics it has. It thus presupposes a given semantics, and cannot be the basis for identifying the semantic values, such as truth conditions, of that language. Others argue, in addition, that Saul Kripke and Hilary Putnam have shown that the foundational semantics of proper names and natural kind terms do not in general involve speakers having any uniquely identifying beliefs about the reference of these expressions, but only their standing in some causal/historical relation to the referents. It follows that knowledge of meaning is not even in general a part of foundational semantics.

From the radically different perspective of fictionalism about knowledge of meaning, these criticisms are not compelling. On this view the phenomenon of language consists fundamentally of speakers' abilities to perform various linguistic actions. The semantics of an actual language in use cannot be read off what is directly observable, the performance of linguistic actions. Rather, the identification of the semantic values of an actual language proceeds via ascription of knowledge of semantic theories to speakers and evaluating their success in explaining linguistic actions. Thus, if ascription of knowledge of some type of truth conditions is explanatorily idle, i.e. knowledge of these truth conditions is not manifestable, then these truth conditions cannot be taken to be the semantic values of sentences of the language in question. More generally, on this view there is no such thing as the prosecution of descriptive semantics independent of a version of foundational semantics.

What about the causal/historical theory of reference? From the perspective of the fictionalist view, the causal theory need have no consequences for the manifestation argument. The significance of the causal theory is, in the first place, that ascription of uniquely identifying beliefs about reference is *not* required to explain use of proper names or natural kind terms. It doesn't follow from this that ascription of knowledge of truth conditions is not required to explain use of sentences in general, including sentences containing these terms. Hence the acceptance of the causal/historical picture doesn't by itself rule out showing that recognition transcendent truth conditions are not

is not an "interior state" just the claim that speakers' conceptions of meaning have no psychological reality? Moreover, consider his challenge to realists to show how knowledge of realist truth conditions can "be viewed as a description of any practical ability of use" (Wright 1993: 17). How is this different from a demand that the ascription of this knowledge must explain some linguistic ability? I do not claim that Wright must be understood in this way; however, it is not at all obvious that Wright's version of the manifestation argument is not fundamentally epistemological in character. On this point I disagree with Miller (2002a: 376 fn. 15).

¹³ See Devitt and Sterelny (1987). However, my account of this objection follows ideas presented in Stalnaker (1997).

the meanings of sentences because ascription of knowledge of such conditions fails to explain our use of sentences.

The second main line of criticism of the manifestation argument applies to both substantial and fictionalist views of knowledge of meaning. The rejection of realist truth conditions rests on a particular account of how knowledge of truth conditions is manifested. Knowledge of truth conditions equips a speaker with the capacity to recognize those conditions as obtaining when they do, and failing to obtain when they do not. The exercise of this recognitional capacity then enables the speaker to make or reject assertions correctly, using the sentence with these truth conditions. But, the objection goes, why should we assume that knowledge of truth conditions results *only* in a capacity to recognize the obtaining of those very conditions? Why could such knowledge not lead to capacities to recognize other sorts of condition, for instance, “those which (given our general theory of the world) constitute evidence, more or less good, for or against the truth of the sentence[, or] even those which point to the unavoidable absence of evidence either way” (Strawson 1977: 16–17)? Alternatively, why could such knowledge not lead to other sorts of capacity, for instance, capacities to interpret others (McGinn 1980)? If knowledge of truth conditions lead to these other capacities, then could we not, on the fictionalist view, explain linguistic performances that depend on those other capacities by the ascription of this knowledge? Alternatively, could we not, on the substantial view, take such performances to be evidence for the speakers’ knowledge of truth conditions?

It should be emphasized that this argument does *not* show that realist truth conditions *are* manifestable after all; it merely shows that the current version of the manifestation argument has *not established* that realist truth conditions are *not manifestable*. A defender of the manifestation argument would have to demonstrate either that capacities mentioned above, distinct from that for the recognition of truth, do not in fact result from grasp of truth conditions, or that none of these capacities result from knowledge of realist truth conditions. In contrast, a defender of realist truth conditions would have to go beyond the present objection, to show not only that these additional capacities result from knowledge of truth conditions, but also that the truth conditions in question are recognition transcendent. The dialectical situation in the wake of this criticism is thus a standoff.¹⁴

I turn now to consider an argument by Wright that may be understood as an attempt to advance beyond the present dialectical standoff. Wright accepts that the abilities

¹⁴ I emphasize this point because philosophers inclined to realism often appear to misconstrue the dialectical situation. For example, in one of the most thoughtful criticisms of the manifestation argument, Miller (2002b) assumes that the capacities identified by Strawson and McGinn obviously result from grasp of realist truth conditions, so that the anti-realist at this point is forced to concede that knowledge of realist truth conditions *is* manifestable. But once this assumption is made clear, one sees that it begs the question against the anti-realist. Given this misconstrual, it is then easy to misread Wright (1993) and Hale (1997) as moving to a weaker version of the manifestation argument which merely attempts to show that the neighborhood abilities *also* result from grasp of recognition immanent truth conditions, and so there’s no basis for preferring the hypothesis that the truth conditions known are recognition transcendent.

identified by Strawson and McGinn do result from knowledge of truth conditions. However, he holds these abilities to be, in some sense, less central to knowledge of truth conditions than the ability to recognize the obtaining of those conditions. Wright argues for this by considering a class of decidable statements:

Imagine that you are seated before a table on which are placed a number of bowls, containing various substances. How might you manifest, in this context, your understanding of statements, like “this one is salty”, “this one is bitter”, “this one is sweet”, etc.? (Wright 1993: 17)

Wright makes two crucial claims about these statements. First, “recognizing the taste of something by tasting it is recognizing that” which makes descriptions of its taste true or false, so, “[t]here would be some point in identifying this range of recognitional abilities with knowledge of the truth-conditions of the relevant statements” (ibid.) Second, although it seems possible that, for example, someone can’t taste sweetness but has the other abilities to use the term “sweet,” such a person “would only questionably be regarded as (fully) understanding the statement that sugar is sweet” (1993: 18).

Let’s call these abilities “central,” and, following Simon Blackburn (1989), the abilities cited by Strawson and McGinn “neighborhood.”

Once these two claims are accepted, it is fairly easy to mount a “challenge” to semantic realism. For the first claim suggests a generalization: the ability to recognize the “truth-conferrers” of any decidable statement just *is* knowledge of its truth conditions. The second claim similarly generalizes to the view that none of the neighborhood abilities with respect to decidable statements are knowledge of their truth conditions. Now, nothing seems to bar our extending these views to all statements, including undecidable ones. But if we do, there seems to be a problem for semantic realism. We don’t know that we have the ability to recognize that the truth conditions of an undecidable statement obtain, or the ability to recognize that these conditions fail to obtain. Semantic realism insists that these conditions nevertheless either do obtain or fail to. So we don’t know that we have the ability to recognize the presence of the truth conferrers or of the falsity conferrers of an undecidable statement. But then it follows, given extension of Wright’s first claim to undecidable statements, that, for all we know, we don’t know the truth conditions of such a statement. That is to say, we lack any justification for thinking that we understand undecidable statements. The challenge, then, is to specify a “practical ability which stands to understanding an evidence-transcendent truth condition as recognitional skills stand to decidable truth-conditions” (Wright 1993: 23).

But the question, of course, is why we should accept Wright’s identification of knowledge of decidable statements’ truth conditions with ability to recognize their obtaining. All Wright tells us is that “there would be some point” in making this identification. But why? Is it just a brute intuition?

I will now argue for Wright’s identification on the basis of what I take to be a clearer intuition. Let’s begin by asking: if someone can for example identify (at least some) things as tigers, and distinguish them from (at least some) things that are not tigers, is there any reason to deny that (to some extent at least) she knows what tigers are? Alternatively,

do we have any reason to deny that she has some conception (perhaps one that she can't fully or clearly articulate) of what qualifies something as a tiger? I think that our intuitive answer is "no" to both these questions. Note well that I'm *not* suggesting that in order to know what tigers are, or to have a conception of what makes something a tiger, someone has to have the capacities just described. It's rather the other way around.¹⁵

I suggest that we have intuitions parallel to these about tigers with respect to more abstract things like conditions. That is, if someone can tell (at least in some cases) that certain conditions have obtained, and can also recognize that those conditions have not obtained, then we have (no) reason (not) to take her to have a conception of what those conditions are, or of what qualifies a set of circumstances as the presence or absence of these conditions. This, I take it, is the intuitive basis of Wright's central claim. Being able to recognize what makes a description of taste true or false is being able to tell whether conditions making that description true are present, and, intuitively, that suffices for knowing what these conditions are; that is, for having a conception of the truth conditions of that description.

More generally, we have an intuition that being able to identify ϕ s, in the sense of recognizing the presence of ϕ s, is sufficient for knowing what ϕ s are or having a conception of what it is to be a ϕ . That is, intuitively, capacities for identification and discrimination of ϕ s suffice for knowing or having a conception of what ϕ s are.

Now we can provide a revised argument for the manifestation challenge. What is our basis for taking a speaker *A* to know what the conditions are in which a taste sentence *S* is true? It is that *A* can recognize the obtaining of certain conditions as ones rendering *S* true, and recognize the obtaining of other conditions as ones rendering *S* false. Now, let's suppose, first, that *A* also has a conception of a condition *C* which makes some other sentence *S'* true if *C* obtains, and, second, that she cannot recognize that *C* has obtained. On the second part of this supposition, *A* obviously cannot recognize the presence of any conditions as making *S'* true. But then we have lost our intuitive basis for taking *A* to have a conception of *C* as conditions that render *S'* true, i.e. for the first part of the supposition. Until we come up with a different basis for taking *A* to have a conception of *C* as a condition for the truth of *S'*, one that doesn't conflict with the second part of our supposition, it's not clear that our two-part supposition is coherent. But the two-part supposition just is the hypothesis that *S'* has recognition transcendent truth conditions. So we don't appear yet to have any grounds for accepting that the notion of recognition transcendent truth conditions is coherent.¹⁶

¹⁵ It seems to me that we do have certain intuitions about the absence of these capacities for identification, but these are not intuitions about individual subjects who lack these capacities. We have the example of Putnam himself as someone who cannot identify any tree as an elm, nor any tree as not an elm, but who we are largely willing to take as having some conception of what elms are. (Although what would we say about a certain Putnam* who is not in a position to distinguish elms from rocks, or cars, or corporations, or tickles, or prime numbers?) What though, if everyone, past, present, and future, in the language community to which Putnam belongs is in the same position as Putnam? Do all or any of members of this community have *some* conception of what elms are?

¹⁶ This construal of the manifestation challenge obviously does not rest on the principle that all features of the meaning that a speaker associates with a sentence must be constituted by facts about

Naturally I don't take this argument to be decisive. In particular it faces two interesting recent critiques of the manifestation argument: Miller's (2002b) revival of an argument due to John McDowell (1981) for the manifestation of realist truth conditions on the basis of the "platitude" that to assert a proposition is to present it as true; and Darragh Byrne's (2005) argument for the manifestability of semantic realism on the basis of compositionality considerations. Here I can do no more than record my conviction that neither is conclusive, and so that the status of the manifestation argument remains an open question.

16.4 WRIGHT'S SEMANTIC APPROACH: CRITIQUE OF DEFLATIONISM

Wright's version of the semantic approach rests on a view of truth he calls minimalism, developed from a critique of deflationist theories of truth. There has been quite a lot of discussion of Wright's argument against deflationism, so I will not here go into the intricacies of how exactly it should be formulated, nor assess its soundness.¹⁷ Instead, I will attempt to sketch a philosophical context for understanding the significance of that argument. I will show, in particular, that Wright conceives of deflationism as including a rejection of (an aspect of) ontological metaphysics, and that Wright's minimalism shares this rejection.

The target of Wright's critique appears to be a fairly standard brand of deflationism. The fundamental claim is that truth is not a property, or not a philosophically interesting or substantial property. Rather, all there is to truth is the fact that a truth predicate is a device of disquotation conforming to the schema

$$"p" \text{ is true iff } p \quad (\text{DS})$$

which can also be used to endorse assertions (Wright 2003: 249–50).

The central premise of Wright's argument is that assertion is a normative practice governed by standards of correctness. So, endorsing an assertion is (re)commending it as meeting *a* standard governing assertoric practice. Now the question is: if one endorses an assertion by using 'true', what standard of assertoric practice does one present that assertion as meeting? Wright claims that a deflationist is committed to denying that the standard could be any norm distinct from being epistemically justified, i.e. being the

her abilities to recognize the obtaining of that sentence's truth conditions. So it is not vulnerable to an argument due to Miller that the manifestation argument is not well-motivated because it rests on this faulty principle.

¹⁷ See Azzouni (ch. 17 in this volume).

expression of a belief that one is warranted as holding. The reason is, “if there were a distinct such norm, it could hardly fail to be reckoned as genuine property of [an assertion] that it did, or did not comply with it” (Wright 2003: 252), but deflationism is committed to truth’s not being a substantial property.

Now Wright argues that the view that ‘true’ “signals” the norm of epistemic warrant is incompatible with (DS), together with the assumption that all sentences in the discourse in question have meaningful negations. Applying (DS) to the negation of a sentence p yields

$$T(-p) \equiv -p$$

But if we negate both sides of (DS), we get

$$-T(p) \equiv -p^{18}$$

Transitivity of the biconditional then takes us to

$$T(-p) \equiv -T(p)$$

If ‘true’ signals being justified, then it seem “not true” should signal not being justified. But then the conclusion just reached says that the negation of a sentence is warranted just in case that sentence is not warranted. This however, fails to hold whenever we are in what Wright terms a “neutral state of information” with respect to p : we don’t have enough information either to justify accepting p , or to justify accepting not- p .

Most of the extensive scrutiny to which this argument has been subject focuses on the question of whether the notion of neutral state of information is itself metaphysically neutral.¹⁹ Here I want to set that issue aside, and raise a much simpler question. Let’s go back to the first stage of the argument, and the conclusion that deflationism must insist that a truth predicate does not “signal” a norm of assertion distinct from current warrant. Does this mean that ‘true’ expresses the *property* of being currently warranted? If so, isn’t deflationism already defeated? Well, perhaps being warranted is not a property of assertions? Why? Perhaps only the version of deflationism which holds that truth is not a property at all is defeated, not the version which holds that truth *is* a property, but not a philosophically substantial or significant one. But that just pushes the question back: why is being currently warranted not a philosophically substantial property? In a footnote Wright tells us that the first part of his argument does not show that deflationism is committed to a truth predicate’s expressing the property of being warranted, since a predicate might express “no property but is used to commend items for their possession of a certain property” (2003: 252 fn. 15). It’s not altogether clear what possibility is described here, but

¹⁸ This transition requires only *reductio ad absurdum* and conditional proof and so is intuitionistically valid.

¹⁹ Tennant (1995); Shapiro and Taschek (1996); Hand (1998).

a different version of the same puzzle arises. Presumably Wright makes this distinction so that deflationism can continue to hold that 'true' expresses no property. But presumably also the "certain property" "signaled" by 'true' is current warrant, and so would be expressed by the predicate "is currently warranted." Now the question is, why the invidious distinction between the two predicates, 'true' and 'currently warranted'?²⁰

The answer is that Wright conceives of his target as a position fundamentally opposed to the following aspect of metaphysical analyses of truth. An account of truth as, for example, correspondence may be taken to furnish a criterion of adequacy for some of our main cognitive and linguistic practices. It is a platitude to say that we try to have true beliefs, or at least to avoid false ones, that we try to make true statements, so long as they are not offensive, or redundant, or misleading. Our practice of asking for justification for what we say or believe seems to be aimed at assuring that our assertions and beliefs are true. The correspondence theory seems to tell us what it is that we're aiming at when we aim at truth, and so explain, in the prior and independent terms of correspondence to reality, what constitutes success for all our diverse ways of justifying our assertions and beliefs. What deflationism, as Wright conceives of it, is fundamentally denying, in holding that truth is not a substantial property, is the existence of such a general account of the basis of successful justification. Hence the invidious distinction between truth and warrant. To claim that truth consists of the property of warrant does not fund an analysis of successful justification in independent terms; it is, rather, merely to advance the (near) tautology that what makes a justification successful is that the assertion or belief in question is currently warranted. The basic picture of deflationism, thus, is that our cognitive and linguistic practices are philosophically primary: they can be described, but not explained or assessed in terms independent of and external to those practices.

Nothing in Wright's inflationary argument goes against this type of deflationist rejection of (ontological) metaphysical analyses of truth. What the argument shows, rather, is that it's a mistake to move from this rejection to the conclusion that in our practice of assertion truth just *is* current warrant. So long as we use a truth predicate to endorse assertions, and have the intuition that it satisfies the disquotation schema, we are committed to conceiving of assertions as governed by two distinct norms, one expressed by current warrant and the other by truth. Note *how* this conclusion is reached: by reflection on features internal to assertoric practice, not on the basis of an independent account of the reality to which that practice relates. Thus, Wright's procedure, in criticizing deflationism, is of a piece with the philosophical picture underlying deflationism.

Wright's minimalism about truth continues this form of philosophical investigation. He begins by pinpointing other intuitive principles about truth and assertion, and arguing for further principles on their basis:

- To assert a proposition is to present it as true, so sincere assertion is expression of belief. (This is sometime called the "assertion-truth platitude.")

²⁰ See Miller (2001) for a related point.

- From the last principle, Wright claims, we derive the equivalence schema,

It is true that p if and only if p ' (ES)

- A sentence is true just in case so is the proposition it expresses. (Wright claims this principle and (ES) together entails (DS).)
- Every assertible sentence or proposition has an assertible negation.
- An assertible sentence " p " says that p .
- The facts are as the assertible sentence " p " says they are if and only if p .
- The facts are as the assertible sentence " p " says they are if and only if p is true. (This follows from the last principle by (DS).)²¹

The last principle leads to a "discourse-internal" interpretation of the correspondence theory of truth if we assume that

- An assertible sentence " p " corresponds to the facts if and only if the facts are as " p " says they are.

Wright continues from these principles about truth and assertion to two claims. First, truth is no more, and no less, than whatever property or norm is signaled or expressed by a predicate that satisfies these intuitive principles. For the second claim we need the idea of an assertoric discourse: a "class of sentences with assertoric content" (Wright 1992: 35), that is, sentences utterances of which are governed by norms of warranted assertibility (Wright 1992: 28). The claim is that for any assertoric discourse it is possible to define a predicate that satisfies the intuitive principles of truth. This is the predicate of superassertibility. In making an assertion, I am of course committed to having a justification for it. Does it make sense to suppose that, at the same time, I also have reasons to think that if I look into the matter further, I will uncover information for overturning my present grounds for making that assertion? To the extent that we find this supposition unintelligible, we take assertion to involve commitment to current justification's survival of additional investigation. This is superassertibility, a norm governing assertion beyond current warrant. Like truth, it coincides in normative force with current warranted assertibility: any reason to endorse an assertion as superassertible is a reason to endorse it as currently warranted, and vice versa. But, Wright argues, if we adopt the principle of evidential constraint for truth of a discourse, i.e. for any sentence p

If p then it is knowable that p , (EC)

then

p is not superassertible \equiv $\neg p$ is superassertible

²¹ Two more controversial ones are: truth is absolute (there is, strictly, no being more or less true), and stable (if a content is ever true, it always is).

also holds. This means that superassertibility diverges in extension from current warrant, exactly like a truth predicate, and so is a “model” of truth for any discourse satisfying the above principle (Wright 1992: 57–61). Superassertibility is a notion of truth explained in terms of standards we acknowledge as applying to our assertions.

Once again we see that Wright’s procedure, in developing minimalism, is of a piece with the deflationist view that assertoric practice is fundamental, to be described but not explained in ontological metaphysical terms. If we take the core commitment of deflationism to be a rejection of ontological metaphysical groundings of the concept of truth, then Wright’s minimalism has the same core commitment, and can be thought of as a continuation of deflationism, albeit a continuation by other means, means that avoid the internal incoherence of deflationism.

16.5 WRIGHT’S SEMANTIC CRITERIA FOR REALISM

The minimalist conception is not an analysis of truth, not an account of what truth consists in for anything truth-apt. It is, rather, an exposition of the marks of the concept of truth in our assertoric practice. A useful analogy is with “surface” indicators of natural-kind concepts such as water. Just as these indicators might pick out distinct stuffs in various possible states of the world,²² so the set of intuitive principles governing truth might be satisfied by distinct properties in various discourses. And, just as the stuffs satisfying the indicators may have distinct underlying chemical structures, so the satisfiers of the platitudes about truth for distinct discourses may have distinct additional characteristics. It is this point that opens a space for prosecuting metaphysics semantically on the basis of minimalism. For, these additional characteristics of truth can be used for explications of forms of objectivity of truth; that is, they can be used to devise criteria for a discourse to be realist. Thus, it might be argued for example that moral sentences may be just as true as sentences about the physical world, but the facts that make them true differ in explanatory potential, and the explanatory potential of moral facts fails to qualify moral truth as objective. It then follows that moral discourse fails to be realist, by this criterion of realism.

Wright advances four criteria for realism, in the form of four “realism-relevant cruces” (2003: 6). The basic outlines are well-known and I will be brief in my exposition. The first criterion is the Dummettian realist view of truth as objective in being recognition transcendent. In Wright’s framework this has the consequence that truth is distinct from superassertibility. Now Wright argues that one limitation of the Dummettian criterion is that in certain discourses, morality for instance, there seems to be room for

²² It should be obvious that this is not to claim that what’s picked out is water in these counterfactual circumstances.

a realist view even though it is fairly plausible that truth is recognition immanent, or, as Wright puts it, evidentially constrained. Thus the second criterion is meant to identify a style of realism with respect to a discourse in which all truths are knowable in principle, so that if a sentence is true, we could, under optimal conditions, arrive at the judgment that it is. The intuitive idea is that such a discourse could still be realist if this coincidence between the facts depicted by our (cognitively) best judgments results from our judgments' tracking, as opposed to constituting, the facts. If so, then it seems that there could be no *a priori* guarantee that best judgment coincides with the facts. The discourse is then said to be Socratic, as opposed to Euthyphronic, when there is an *a priori* guarantee for the coincidence, because the facts are what they are only as a result of our best judgment.

The next two criteria arise from considering how a realist view of truth as correspondence to facts differs from the minimalist interpretation of correspondence mentioned in the last section. The third criterion attempts to capture a realist conception of correspondence. The intuition here is an analogy between using a class of statements purely to represent reality and using a camera to take photographs: "If we take photographs of one and the same scene which somehow turn out to represent it in incompatible ways, there has to have been some kind of shortcoming in the function of one (or both) of the cameras, or in the way it was used" (Wright 2001: 55). Thus truth of a discourse is objective in this sense if it holds *a priori* that differences of opinion over the truth of a claim will involve a cognitive shortcoming in one or the parties to the dispute (unless one of three excusing conditions hold: vagueness in the statement, or in standards of assessment, or variation among people in evidential thresholds). The discourse is then said to exert *cognitive command*. The final criterion attempts to capture a realist conception of facts (Wright 1992: ch. 6). The underlying intuition, from certain moral anti-realisms, is that facts about the physical world are more robustly a part of reality than are moral facts. Physical facts explain not only our true beliefs about them, but also the obtaining of other non-cognitive states of affairs through their direct causal effects, not involving our beliefs. Moral facts, by contrast, explain our beliefs and other attitudes toward them, but can explain other states of affairs only through our attitudes, not via unmediated effects on those states of affairs. Physical facts are, in this way, said to have wide *cosmological role*.

In the rest of this section I will consider how these criteria are mutually related, and how they are related to Dummett's criteria.

We have already seen that the Euthyphro contrast holds only if semantic anti-realism holds. If some sentences can be true undetectably, then in those circumstances the states of affairs they depict would obtain even though we could not be in a position to judge that they do, i.e. we could not constitute those facts by our judgment. Semantic realist discourse cannot be Euthyphronic.

Prima facie the cosmological role criterion cuts across the semantic realism criterion. If there are undetectable facts, then they obtain without our believing that they do. But then those facts would not explain our beliefs that they do. So these facts could not explain any states of affairs via their effects on our beliefs. Hence they cannot have narrow

cosmological role. But semantic realism is consistent with the existence of *detectable* truth. So it seems only a semantic realism that denies the existence of *any* in principle recognizable truths in a discourse rules out facts with narrow cosmological role.

What about cognitive command? Much here depends on what counts as a cognitive shortcoming. If merely being wrong, and not being able to tell who's right, are both cognitive shortcomings, then it seems that semantic realism about a discourse implies the exertion of cognitive command. If so, then cognitive command can fail only in semantic anti-realist discourses.

The remaining question about how the three explicitly non-Dummettian criteria are related to the Dummettian one is whether cognitive command can fail in a semantic anti-realist discourse. It's clear that if cognitive command cannot fail in such a discourse, then it can't fail in any minimal discourse. But then all minimal discourses would be realist by this criterion, and then, from Wright's perspective, cognitive command would be "trivialized." Taschek and Shapiro give a much discussed argument for this conclusion. Since the discourse is, by assumption, semantic anti-realist, it obeys the evidential constraint that all truths are knowable. Suppose now that there is a disagreement—*A* believes *P* and *B* believes not-*P*, and suppose for *reductio* that cognitive command fails, so neither *A* nor *B* has a cognitive shortcoming. Suppose in addition that *P* is true. By (EC), *P* is knowable. So *B* could have come to believe, correctly, that *P*. But then her belief that not-*P* indicates that she didn't do as well as she could have in forming true beliefs. Surely this means *B* has a cognitive shortcoming, contradicting the assumption that neither she nor *A* has a cognitive shortcoming. So *P* is not true. Note that at this step we have discharged the temporary assumption that *P* is true, so the conclusion that *P* is not true now depends just on the initial assumption of a cognitively blameless disagreement. Since the discourse is minimal, not-*P* is true. By (EC), not-*P* is knowable. So *A* could have come to believe, correctly, not-*P*. But then his belief that *P* is a cognitive shortcoming, again contradicting the assumed absence of such shortcomings. Hence the assumption that *A* and *B* disagree is inconsistent with the assumption that neither has a cognitive shortcoming.

Wright's answer to this argument begins by claiming that the presence of cognitive command in this case is the existential or disjunctive thesis that

Either *A* has a cognitive shortcoming or *B* has a cognitive shortcoming

What the argument shows is that the assumption that *A* disagrees with *B* over *P* is incompatible with the negation of this disjunctive claim. So, what is immediately implied by *A*'s disagreeing with *B* is the double negation of the disjunction, and classical reasoning seems required to reach cognitive command. It follows that Taschek and Shapiro's argument fails if one can rule out the use of classical reasoning. If so, we can take their argument to rule out that neither *A* nor *B* has a cognitive shortcoming, without demonstrating that one or the other has such.

Wright then argues that disagreements which, intuitively, don't exert cognitive command involve statements for which classical reasoning fail. These "disputes of

inclination”—over e.g. whether snails are delicious—“might persist even though there seemed to be nothing else relevant to it about which either party was ignorant or mistaken” (Wright 2001: 47). At this point our intuition is that “the disputants should just ‘agree to differ’, as it were, without imputation of fault on either side,” because the opinions here derive from differences in “personal inclination,” and don’t “have to answer to . . . the facts” (ibid.) The basic problem raised by disputes of inclination for minimalism is that the sentences involved seem minimally assertoric. Hence there is a minimal notion of fact, as whatever renders these truth-apt sentences true. So it seems that these disputes have to answer to the facts after all, and the side of the dispute which gets the facts wrong is at least mistaken, if not cognitively deficient in some stronger sense.

Wright’s solution is to accept that these disputes concern minimal facts, but claim that it is indeterminate what the facts are. The indeterminacy, as we will see, is explained epistemically. The statements at issue in a dispute of inclination present what Wright calls “quandaries.”²³ This means that we have four levels of ignorance with respect to such a statement q ; we:

- don’t know whether q is true,
- don’t know how we might come to know whether q is true,
- have no reason to think that there is any way of coming to knowing whether q is true, and
- have no reason to think that it is even metaphysically possible to know whether q is true. (Wright 2001: 71–5)

Now, the double negation of the law of excluded middle is provable in minimal logic. Hence we have logical grounds for ruling out the negation of any instance of it for a quandary. But do we have grounds for accepting that instance? Suppose we do. Then, since (EC) holds, we would have grounds for accepting that

Either it is feasible to know that q or it is feasible to know that not- q

But, given the four levels of ignorance with respect to q we don’t know that we can either recognize that q holds or recognize that its negation does. So we don’t know that we have grounds for accepting any quandary instance of excluded middle. Hence eliminating the double negation of such an instance would take us from a statement for which we have a justification to one for which, for all we know, we cannot have a justification. This is the argument against classical reasoning for quandaries.

The situation for a quandary q is that we have grounds for ruling out that the world is not rightly described by either q or not- q . This is just to say that, since q is minimally

²³ The theory of quandaries is also applied to two other issues: an intuitionistic solution to the Sorites, and more generally the rejection of classical logic on the basis of semantic anti-realism.

assertoric, it can't be that the world fails to serve up any fact rightly or wrongly described by *q*. But we *don't know* that we can recognize *which* fact is served up by the world. That is to say, it is (epistemically) indeterminate what the facts relevant to *q*'s truth-value are.

One worry here is that Wright's solution fails to capture the original intuition about disputes of inclination: it cannot be ruled out a priori that neither of two disagreeing parties is wrong. On the quandary view, it *can* indeed be ruled out in every disagreement that no one is wrong. Of course it's indeterminate which party is wrong, in the sense that we don't know we can find out which. But is this enough to sustain the intuition that we should just agree to differ?²⁴

This discussion of cognitive command suggests that it's a descendent of Dummett's determinacy criterion for realism. There are, of course, significant differences. Dummettian indeterminacy lies in the absence of uniqueness in truth fixation: either there is no unique state of affairs that obtains, or there is no unique state of affairs represented. Wright's indeterminacy is epistemic: absence of unique truth fixation can be ruled out, but we don't know we have epistemic access to it. Because Dummett's criterion is non-epistemic, it cuts across the semantic realism/anti-realism divide. As the example of neutralism shows, a semantic realist with respect to an area of discourse can nevertheless conceive of it anti-realistically, in Dummett's sense, as being indeterminate.

The final point I want to note is that Wright explicitly connects the wide cosmological role criterion to Dummett's view of abstract objects (Wright 1992: 178–82). As we have seen in section 16.2, Dummett holds that the ascription of reference to a singular term may rest on no more than the satisfaction of a set of logico-syntactic features; put in Wright's terms, Dummett holds a minimalist conception of reference. Just as, in Wright's scheme, realism consists of ascribing more-than-minimal features to truth, so realism about reference can be understood as the claim that singular terms in a discourse have more than the minimal features of reference. What I have argued above about Dummett's classical semantic criterion of realism can be formulated as the claim that for a term to refer realistically, its referent must be an aspect of the truth conditions of sentences containing that term. This fails in the anti-realist neo-Fregean semantics for direction terms. Wide cosmological role is obviously a distinct more-than-minimal feature; for one thing it applies to facts rather than referents. But the underlying idea is the same: for reference, or fact, to be genuinely a part of reality it has to be more than merely a reflection of our discursive practices.²⁵

²⁴ In addition, in Shieh (forthcoming) I argue that if classical reasoning is allowed, then Wright's four levels of ignorance with respect to quandaries don't suffice to establish that excluded middle is not known to apply to quandaries.

²⁵ There is a much murkier set of questions about how Wright's "non-Dummettian" criteria are related to one another in a semantic anti-realist discourse. Here is one instance. How are cognitive command and the Euthyphro contrast related? If a discourse is Euthyphronic, does it exert cognitive command? The assumption here is that truth is constituted by best opinion. Suppose there is a disagreement. Must

16.6 META-PHILOSOPHICAL CONSIDERATIONS ON THE SEMANTIC APPROACH

Now that we have seen some conclusions and arguments distinctive of semantic metaphysics, we turn to meta-philosophical questions about this approach. To begin with, why pursue this approach? Dummett (1963) initially proposes it as an alternative to the traditional ontological conception of metaphysical disputes outlined in section 16.1. He lists three reasons for pursuing this alternative. First, in certain disputes, such as those over the reality of the past and the future, there isn't a class of entities the existence of which divides realism from anti-realism. Second, there seem to be a number of analogies among traditional disputes, which can be brought out by formulating them in non-ontological terms. Finally, in the case of mathematics, the existence of entities "seems to . . . deflect the dispute from what it is *really* concerned with" (Dummett 1963: 146; emphasis added). Dummett notes, in addition, that his approach does not exclude treatment of questions of the existence of entities, but interprets them as issues about statements purporting to refer to those entities (1963: 147); although, as we'll soon see, he admits that this maneuver fails to capture all traditional ontological issues. Subsequently Dummett adds another consideration (1982; 1993). As we have seen, one central traditional form of opposition to realism is reductionism. A classic example is behaviorism, which denies that there are any mental facts not reducible to complex facts about behavior. Reductionist positions, when prosecuted by attempting to translate, for example, statements about mental states into statements about behavior, have run into decisive problems. But the semantic construal of realism shows that anti-realism doesn't have to rest on such problematic translatability claims, and so opens up new possibilities for metaphysical debate.

All this suggests two versions of Dummett's research program, one modest and the other more ambitious in its aims. The modest view presents the program as just one way of doing metaphysics among others, promising illumination on traditional issues. The stronger view presents the program as revealing what's genuinely at stake in metaphysical disputes.

It's plausible that Dummett's considered view is a modest one, for he specifically notes a number of limitations of his program. It does not apply to the traditional debate between realism and nominalism about universals, because nominalists deny that general terms refer, not that statements containing them are recognition transcendent

at least one party have a cognitive shortcoming? One might argue as follows. Under the Euthyphronic assumption, contradictory best opinions would imply true contradictions. If we reject that possibility, we seem to have to conclude that no disagreement results from contradictory best opinions. So one opinion must be less than best, and does that not mean one party has a cognitive shortcoming? So it seems Euthyphronic discourses must exert cognitive command.

(Dummett 1963: 147). Medieval nominalism, of course, is one salient ancestor of the modern error theories of Mackie and Field. Another limitation is that the program does not apply to the dispute between expressivism, “which den[ies] the status of statement to . . . moral utterances,” taking them rather to “express *attitudes* incapable of objective justification,” and the type of moral realism which holds these utterances to be “statements [that] can, in favourable circumstances, be objectively established as true” (Dummett 1993: 466–7).

In this section I will argue for a version of the modest view as the most philosophically interesting way of thinking of the semantic approach. But, to reach that conclusion, I begin by considering an influential account of the ambitious version and a set of criticisms based on it due to Devitt and Miller. On this reading, Dummett claims that his program is the only legitimate way of doing metaphysics, on the basis of two theses. The “metaphor thesis” (Devitt 1983: 80) is that any ontological formulation of realism and anti-realism is merely metaphorical. The “constitution thesis” (Miller 2005) is that the non-metaphorical, literal, contents of realisms are conceptions of meaning, specifically, the recognition transcendence of truth.

One key basis for taking Dummett to hold the metaphor thesis is this passage:

[F]or platonists, [mathematical statements] are about abstract structures, existing independently of our knowledge of them; for intuitionists, they are about the free products of human thought.

[W]e have here two metaphors: the platonist compares the mathematician with the astronomer, the geographer or the explorer, the intuitionist compares him with the sculptor or the imaginative writer . . . (Dummett 1978: xxv)

It’s *prima facie* unclear, and Dummett never explains, why the platonist’s apparent claim that mathematical objects exist independently of our knowledge is a metaphor comparing the mathematician to the astronomer, or why the intuitionist’s apparent claim that mathematical objects are creations of human thought is an equally metaphorical comparison of the mathematician to the sculptor.

Devitt, it seems, finds only one argument by Dummett that might bear on the metaphor thesis: “[a]ssuming a platonist ontology does not lead to a platonist view of meaning and logic; assuming an intuitionist ontology does not lead to an intuitionist view of meaning and logic” (1983: 81). He points out that this conclusion doesn’t establish the metaphor thesis, since “[t]he mere fact that a disagreement does not have any consequence for semantics does not show that it is not a real disagreement, does not show that it is only metaphorical” (ibid.) Of course, perhaps Dummett holds that what qualifies an ontological thesis as metaphorical is that it has no implications for semantics. But then Dummett’s justification of his program is obviously flawed. It is supposed to be because ontological theses are metaphorical that one should conduct the debate in meaning-theoretic terms. So, it is question-begging to argue that the reason why opposed ontological theses are mere metaphors is that they do not imply distinct theories of meaning.

Miller points out another problem for the metaphor thesis. Presumably Dummett doesn't think that the recognition transcendence of truth is a merely metaphorical thesis. But, one can use the notion of recognition transcendence to formulate a sharpened version of Devitt's (ontological) "Common-Sense Realism" (1991: 24). That is, instead of:

Tokens of most current observable common-sense and scientific physical types exist, and the fact that they exist and have properties such as mass, size, shape, color, and so on, is independent of anyone's beliefs, linguistic practices, conceptual schemes, and so on,

we can replace independence of beliefs, practices, etc. with recognition transcendence:

Tokens of most current observable common-sense and scientific physical types exist, and in general there is no guarantee that we will be able, even in principle, to recognise the fact that they exist and have properties such as mass, size, shape, colour, and so on. (Miller 2005: section 6; 2006: 987)

If the claim that truth conditions are recognition transcendent is not metaphorical, why is this revised version of common sense realism so? The upshot of these objections is that the metaphor thesis is unclear or unjustified, and in any case provides no independent support for the constitution thesis.

Devitt further argues that the constitution thesis is false, because it leads to a counterintuitive classification of sense-data idealism as a variety of realism: "An idealist who believed in the objective existence of a purely mental realm of sense data could . . . believe that physical statements are true or false according as they do or do not correspond to the realm of sense data, whatever anyone's opinion on the matter" (1983: 77).

I turn now to argue that neither Devitt and Miller's reading of Dummett nor their criticisms is fully compelling. I begin with Devitt's objection to the constitution thesis. There's no doubt that it's counterintuitive to classify such forms of idealism as realism. But what exactly makes it counterintuitive? I will now show, using the case of Berkeley's idealism mentioned in section 16.1, that what makes this classification intuitively wrong has nothing to do with the semantic approach.

Devitt's view of common-sense realism is course specifically directed at physical objects, but it would surely be a point against it if it rules out ontological realism about mental entities and their properties. How would one formulate such a type of ontological realism about the mental? Let's note, to begin with, that Miller's sharpened formulation of common-sense ontological realism is a conjunctive claim: physical tokens exist objectively, and, their existence is recognition transcendent. In section 16.1 we saw that one form of realism asserts independent existence in the sense of objective existence: that is, the entities in question exist no matter whether they are taken to exist by us. It seems implausible that Miller's phrase "objectively exist" means this concept of objective

existence, since then his formulation amounts to the claims that physical tokens exist independently of being taken to exist, and independent of being recognized to exist, but in this context recognizing that something is the case surely amounts to correctly taking it to be the case. My conjecture, then, is that Miller's objective existence is the other sense of independent existence of section 16.1: fundamental or irreducible existence. Moreover, it's clear that the reduction being denied is reduction to the mental (see Miller 2006: 987). So, Devitt's ontological realism about physical objects, as sharpened by Miller, is the claim that physical tokens exist, are not reducible to mental items, and the fact that they so exist (and have the properties they have) is recognition transcendent. Let's now try to transpose this formulation to ontological realism about the mental. The result, surely, is:

Tokens of most mental (or psychological) types exist, are not reducible to mental entities, and in general there is no guarantee that we will be able, even in principle, to recognize the fact that they exist.

It's not clear that the first two conjuncts of this claim together make any sense, since they seem to imply that tokens of mental types are not reducible to mental entities. So ontological realism about the mental would have to drop the first two conjuncts, and be something like: it is a fact that mental entities and states exist, with various properties, no matter whether anyone believes or knows or is capable of knowing this to be a fact. In short, a coherent ontological realism about the mental has to be, in my terms, a claim of objective existence, not a claim of fundamental or irreducible existence.

Now let's consider a special instance of common-sense realism: tables exist and have properties such as colors independently of anyone's beliefs. One venerable theory of colors is that they are powers (i.e. dispositions) of objects to cause certain sensations in perceivers under various circumstances. Is this view a realist theory of colors, or an idealist one? Are colors really there, in the objects, or are they projections of our sensory responses onto the objects? *Prima facie*, one could go either way on this question. On the one hand, one might argue that color properties are response-dependent and so no physical objects have colors independent of our mental states. On the other hand, one might argue that the fact that someone responds to an object (in such-and-such conditions) by entering into certain mental states is completely objective: it would be a fact no matter whether anyone recognized it. Is this second view a realist theory of color? If so, what makes the view realist is that, for example, for a table to be brown is an objective aspect of reality, specifically, an objective fact about the relation of the table to our mental states.²⁶

On Berkeley's view, that this table exists is a fact about God's mental states. Moreover, for Berkeley it is an objective fact about God's mental states, one which holds no matter whether anyone recognizes it. So, if it's at least intelligible to be ontological realist about

²⁶ Compare Rosen (1994: section V).

minds and mental states, in the sense sketched above, and if it's at least intelligible to be ontological realist about response-dependent properties such as colors, again in the sense sketched above, why isn't it intelligible to take this Berkeleyan view to be an ontological realist view about the existence of tables? That is to say, even if the constitution thesis is false, Berkeleyan idealism would count as one variety of ontological realism.

Let's approach this conclusion in a different way, using the distinction between two types of independent existence outlined in section 16.1. Recall from that discussion that Berkeley's idealism can be understood as denying that ordinary physical objects exist objectively, or as denying that such objects exist irreducibly. Recall, moreover, that entities reducible to other types of entity may nevertheless exist objectively, if entities of those other types exist objectively. From this perspective, one can diagnose Devitt's objection as stemming from confusing these two types of independent existence. The confusion is natural in the case of Berkeleyan idealism, precisely because the reduction base consists of mental states. But once one realizes that there is an intuitive form of ontological realism about the mental consisting in the thesis that mental states exist objectively, one can see that claims about reductions to the mental may also count as forms of ontological realism.

What this shows is that what is intuitively wrong about taking Berkeleyan idealism to be a form of realism is *not* that the mental states which, on this view, constitute physical objects cannot exist as *objectively* as we intuitively take physical objects to exist objectively. To the extent, then, that ontological realism consists of some claim of objective existence, there is nothing problematic about taking Berkeleyan idealism to be a variety of ontological realism. So why do we think that Berkeley's views is not realism? The basic reason is that Berkeley's view flies in the face of the common-sense (stone-kicking) intuition that physical objects simply are not ideas. That is to say, we intuitively think that physical objects are fundamental aspects of reality, not reducible to other kinds of entities. Since Berkeley denies to physical objects such irreducible or fundamental existence, we take him to hold that the physical world is less than fully a part of reality. Accordingly only an ontological realism based on irreducible existence would rule out Berkeleyan idealism as realist.

This conclusion would still yield an objection against the semantic approach if it cannot account for our intuitive classification of Berkeleyan idealism as anti-realist. Now, semantically Berkeleyan or sense-data idealism is explicated by phenomenalism, a reductive account of the truth conditions of physical object statements in terms of sense-data statements. Dummett's account of phenomenalism is quite complex and in the following I simplify. It is well known that if phenomenalism is to allow material-object statements to be true even though not actually verified by us, then, short of resorting to the Berkeleyan expedient of verification by God, it has to explain the truth conditions of such statements using subjunctive conditional statements. The "prototype" of such an account is "the rendering of 'There is a table in the next room' by 'If anyone were to go into the next room (and switch on the light), he would see a table'" (Dummett 1982: 248). The statement would be false if the "opposite" subjunctive conditional held: "If anyone were to go into the next room (and switch on the light), he would *not* see a table."

By Dummett's criteria, phenomenalism would qualify as a fully realist semantic theory only if exactly one of each of these pairs of opposite subjunctive conditionals is determinately true. But it is implausible that subjunctive conditionals in general satisfy this condition. For example, neither "if I had reached the station after 7, I would have missed the 6:57 train" nor "if I had reached the station after 7, I would not have missed the 6:57 train" is determinately true since whether I miss that train upon arriving at 7 depends on whether it had arrived on time. So what reason would we have to think that the opposite subjunctives involved in the phenomenalist reduction satisfy this condition? A semantic realist about physical object statements holds that, for example, "'There are living organisms on some planet in the Andromeda galaxy' must be determinately true or false" (Dummett 1982: 250). So, she can infer that if this statement is true, so is the subjunctive, "If we were to travel to the Andromeda galaxy and inspect all the planets in it, we should observe at least one on which there were living organisms," and, if it is false, so is the opposite subjunctive. So, for a realist about material object statements, the determinate truth-values of these statements fix exactly one of each of the phenomenalist's opposite subjunctives as determinately true. But a phenomenalist thinks that the truth or falsity of material object statements is determined by subjunctives, so she can't avail herself of this explanation. So phenomenalism has no good grounds for holding that its semantics for material-object statements is fully realist.

I turn now to the metaphor thesis. There is no doubt that Dummett repeatedly characterizes ontological theses as "metaphors" (1978: xxv, 229) and "pictures" (1977: 383; 1991: 10). If these characterizations are assertions, then I agree with Devitt and Miller that Dummett has done little to explain these claims or justify them. But we need not take these characterizations to be assertions. I will now argue that a central passage in the Preface of *Truth and other Enigmas* (Dummett 1978) can be understood as *proposing*, rather than asserting, that ontological theses be construed as metaphors or pictures. On this view Dummett does not hold a metaphor *thesis*.

The text I have in mind begins with the claim, quoted above, that platonists think mathematical statements are about mind-independent entities while intuitionists think they're about products of human thought. It's important to see that the philosophical context in which these contrasting ontological views are advanced is the debate between intuitionists and classical mathematicians over whether the law of excluded middle is valid in mathematical reasoning. The original intuitionist, L. E. J. Brouwer, thought of his rejection of excluded middle as based on the ontological claim that mathematical objects are free creations of the human mind, and took the acceptance of excluded middle by classical mathematicians to rest on belief in an erroneous platonist mathematical ontology of mind-independent entities.²⁷ On this view the debate over excluded middle is to be settled by the ontology of mathematics.

Dummett then describes two problems for this view of the debate. The first is, "How are we to make the disagreement [over the ontological status of mathematical objects]

²⁷ For a canonical statement of these views see Brouwer (1975 [1908]).

into a definite one, and how can we then resolve it?" (1978: xxv). The second is, "What bearing does either metaphysical view about the status of mathematical objects have on the forms of reasoning employed within mathematics?" (1978: xxv–vi). Dummett doesn't address the first problem at all, but in response to the second, he argues that the opposed ontological views of platonism and intuitionism have no bearing on mathematical reasoning, because they do not imply opposing accounts of the truth conditions of mathematical statements. You can take mathematical objects to be mind-independent and yet take mathematical statements to be true only when recognized by us to be true; you can take mathematical objects to be creations of the human mind, and yet think that mathematical statements can be true of those creations even if not recognized by us to be true. Since Dummett holds that truth conditions are central to meaning, he states this conclusion as: "the alternative metaphysical views about mathematical objects do not serve as premises for the rival interpretations of the meanings of mathematical statements, and for the differing consequential views about the validity of logical laws as applied to them" (1978: xxviii).

For our purposes, I will not go into the details of the argument or assess its cogency. Let's instead read what Dummett writes next:

It is more tempting to suppose that there is a dependence in the opposite direction. If one believes, with the platonists, that . . . our mathematical statements [are] all determinately either true or false independently of our knowledge, then one will find it natural to adopt the picture of a mathematical reality existing, fully determinate, independently of us. If, on the other hand, one believes, with the intuitionists, that the content of a mathematical statement resides entirely in our ability to recognise what constitutes a proof of it and what a disproof, so that, when we lack an effective means of arriving at a proof or a disproof, we have no right to declare it either true or false, one will prefer a picture according to which mathematical reality is constructed by us, or, at least, comes into existence only as we become aware of it. (ibid.; emphasis added)

Why does Dummett not simply say that metaphysical views depend on conceptions of truth conditions? Why the tentativeness of "more tempting to suppose"? I suggest that Dummett is here advancing a proposal for explicating ontological theses. Why adopt this proposal? Well, as we saw, Brouwer and his opponents think that who wins the debate over excluded middle depends on whose mathematical ontology is correct. But it's not clear that there are determinate shared standards for adjudicating ontological claims. Moreover, in any event there is a case against the relevance of ontology to settling the debate. Suppose, however, that we take the disagreement over excluded middle to depend, not on whose ontology, but on whose view of the truth conditions of mathematical statements is correct. Then, first of all, there are determinate criteria for the correctness of an account of truth conditions, namely whether it provides a model of linguistic understanding adequate to account for the possibility of communication. Second, these divergent views of truth conditions imply opposed positions on the validity of excluded middle. Finally, since these opposed conceptions of truth naturally

suggest opposed ontological views, the disagreement over truth conditions is a reasonable replacement for the disagreement over ontology, one that captures at least one aspect of what Brouwer and his opponents took themselves to be debating.

Does this account generalize beyond this case of mathematical ontology? One respect in which it differs from the other ontological disputes we've mentioned is that in those cases there are no standing disagreements over logical principles that are claimed to rest on differences of ontology. What that means is that one would not be able to claim, as an advantage of explicating the ontological dispute semantically, that the semantic disagreement is directly relevant to a disagreement over logic while the ontological one is not. But there's no reason to think that the other two grounds for recommending the semantic approach lapse.

This reading suggests that what's really bothering Dummett about ontological theses, that makes him call them metaphors or pictures, is just that he doesn't see determinate standards of correctness for ontological claims. This suggests an answer to Miller's question as to why the formulation of ontological realism in terms of recognition transcendence of existence is any more metaphorical than semantic realism. The correctness of this formulation of ontological realism with respect to a class of entities clearly turns on whether the fact that these entities exist and have the properties we take them to have can obtain independently of our being in principle capable of recognizing that it obtains. But on what basis do we determine whether this is the case or not? If we suppose that a fact obtains just in case a statement depicting it is true, then the correctness of this ontological realism turns on whether the truth of statements about the entities in question are recognition transcendent. And, then, a semantic metaphysician would claim, since we can evaluate whether the truth of a class of statements is recognition transcendent or not on the basis of whether the supposition of recognition transcendence yields an adequate model of meaning, we have a determinate criterion for evaluating the recognition transcendence style of ontological realism. That is to say, if we explicate ontological realism by semantic realism, then ontological realism is not metaphorical. Without this explication, it's an open question whether there are determinate standard for evaluating the truth of ontological realism; that is, whether ontological realism is metaphorical. So the answer to Miller's question is that we have no grounds for taking his version of ontological metaphysics to be non-metaphorical if it is not explicated semantically.²⁸

The upshot of these considerations is a modest version of the semantic approach: it's one way of doing metaphysics among others, but it has the advantage that, formulated semantically, metaphysical questions and theses are subject to determinate criteria of evaluation.

I suspect that traditional ontological realists such as Devitt would find even this modest version of semantic metaphysics troubling. The problem is that the semantic approach seems to presuppose an anti-realist construal of ontology. Ontological

²⁸ It should be clear that this argument does *not* depend on claiming that the notion of fact cannot be explained independently of the notion of true linguistic representation; in this respect it differs from the main argument of Taylor (1987).

categories, the traditional realist would think, are (very general) features of reality. By adopting the semantic proposal, we are, it seems, forced to take these categories to be at bottom just features of our thought or discourse, or, at best, as accessible only through features of our thought or discourse. So, doesn't the semantic approach assume that fundamental general features of reality are *not* independent of our thought or talk? Why isn't that just to hold that ontological categories are *not* objective features of reality? Indeed, one doesn't have to be a realist about ontology to find this problematic. Shouldn't ontology begin from a level playing field that doesn't rule out realism *ab initio*?

Let's consider what exactly the ontological metaphysician takes the semantic proposal to have invidiously ruled out. It is, obviously, that ontological categories are independent of semantic categories. How is this independence to be made out? How about modally? There could be ontological categories in reality distinct from our semantic categories, and any of our semantic categories can fail to be instantiated in reality. And what is the modality in question? How about conceivability?

It would be rash to attempt a fully adequate evaluation of this objection, so let's merely consider the ontological category of fact, a category that, as we saw, figures centrally in the notion of objective existence. The semantic metaphysician suggests that the notion of fact be explained in terms of fulfilled truth conditions, and it then seems to follow that there are no facts in reality distinct from conditions that make our statements true. The ontological metaphysician, in contrast, insists on two possibilities. First, there might be facts in reality that are not fulfilled truth conditions of our statements or thoughts. Second, large classes of what we take to be statements or thoughts might not depict facts in reality. Can we conceive of such facts that exceed our representational grasp?

I will discuss the first possibility drawing on Dummett's characterization of Frege as an ontologist (1981a: 430–2). In one sense we obviously can conceive it: there are statements we have not in fact already made, and surely some of these are conceivably true; there are statements we cannot in fact make, because of our present physical limitations—they are, say, too long—and surely some of those are true; there are, perhaps, some statements we cannot in fact make, because of our present intellectual limitations—they contain, say, words expressing concepts we can't presently grasp—and some of those are conceivably true. So let our semantic metaphysician regroup and reformulate her view as: there are no facts in reality distinct from the obtaining of conditions under which statements that we can in principle understand are true. But why can't there be facts distinct from *these* fulfilled truth conditions? It seems clear we can conceive of the existence of such facts. But can we, in principle, describe such facts? If so wouldn't they render those descriptions, which presumably we can in principle understand, true? So, we can conceive *that* there are such facts, but we can't conceive *what* they are. If we can't, even in principle, form any conception of such facts, what difference could the possibility of their existence make to our metaphysical theorizing? More importantly, if any fact that we can conceive has to be the truth-maker of a statement we can grasp, then a feature possessed by any possible truth condition of a statement will be a feature of any fact we can conceive. There may *be* inconceivable facts, but, being inconceivable, we can have no grounds either for or against their having the same features as conceivable facts.

It is then not clear that whatever facts the semantic approach might miss would make a significant difference to metaphysical theorizing.

The second possibility takes us to Wright's version of the semantic approach. Wright's minimalism, as we saw in section 16.4, rests on taking our practice of assertion at face value, as not requiring any ontological foundations. This is challenged by the second possibility, that some of what we take to be statements about the world don't in fact depict reality. One way this could come about is if we take the purported statements to be about certain entities or properties, but there are no such entities or properties in reality. As I mentioned in section 16.1, there are then three options. One might take all these statements to be false. One might take them to have no truth-value. Or, one might take them not to be genuine statements, not capable of being true or false, at all. Another way this could come about is if we take the facts described by the purported statements to have certain features—for example, they necessitate the obtaining of other facts, or they motivate us, independent of our desires, to certain actions—but no facts in reality could have such features. Here there is usually just one option: the utterances in question are not statements assessable as true or false, and whatever mental states they express are not beliefs about the world. All these options can be taken to be varieties of error theory, based on various ways in which we could be under a misapprehension about the ontological presuppositions of our purported representations of reality. Some philosophers who find us suffering from such illusions about our words want to offer some account of our attachment to the use of these words. One type of explanation is that we don't in fact use these words to represent states of the world, but rather to express various attitudes we have toward those states. This, of course, is expressivism. Thus, error-theoretic and expressivist anti-realisms are not just ontological anti-realisms that don't fit in Dummett's picture of realism and anti-realism. They represent a threat to the entire semantic approach based on minimalism.

With respect to any particular discourse, an expressivist could deny that it is genuinely assertoric, and so not genuinely apt for minimal truth. In response to expressivism Wright adopts a minimal conception of truth-aptitude, i.e. of what qualifies utterances to be assertions assessable as true or false. Only two features are required for a class of sentences to be part of an assertoric practice. First, it has to be closed under certain "syntactic" operations: if a sentence *S* belongs to the class so do the negation of *S*, conditionals in which *S* occurs as antecedent, propositional attitude reports in which *S* occurs as the content clause, and so on. Second, the use of these sentences has to be governed by "communally acknowledged standards of proper use" or of "warrant"; Wright often calls this "discipline." Two central examples subject to expressivist treatment are moral and modal discourse, but these surely have the first of the two features, and plausibly have some of the second, at least relative to certain communities.

It's fairly clear that minimal assertion is by itself not enough of an answer to expressivism. The reason is that the most plausible motivation for expressivism with respect to a discourse is the failure of purported ontological presuppositions for that discourse. For example, why should one think that declarative sentences containing moral vocabulary are expressions of attitudes rather than statements of facts? An obviously

compelling reason is that there are no genuine facts to be described by these sentences, so whatever it is we're doing by uttering such sentences, it cannot be making assertions expressing beliefs about the world. Given Wright's platitude that sincere assertion is the expression of belief, it follows that utterances of moral sentences are not assertions.²⁹ That is to say, an ontological metaphysician's answer to Wright's notion of minimal truth-aptitude is that since truth-aptitude has ontological presuppositions, an error-theoretic view of any minimally truth-apt discourse cannot be ruled out. In response to this maneuver Wright advances a minimalist view of belief as whatever is expressed by the sincere endorsement of a sentence belonging to a minimal assertoric discourse. But this just raises the question, why should we buy the discourse-internal, as opposed to the ontological, conception of belief?

The fundamental issue, then, is minimalism's reply to error theories. Wright's reply begins with an interpretation of error theory from the perspective of minimalism. Within the minimalist framework, there are two possibilities about truth: either it's superassertible or it isn't. In the first case, error theory consists of the thesis that none or few of the assertions of a discourse are superassertible. Now a sentence *p* is superassertible just in case *p* is currently warranted and *p* will not lose its warrant upon further investigation. So, *p* is not superassertible just in case either *p* is not currently warranted or *p* is currently warranted but will lose its warrant upon further investigation. But, as noted above, it's not coherent to take oneself to be currently justified if one has grounds for thinking that the justification will be lost. So, grounds for thinking that justification will be lost are grounds against current warrant. Thus, any argument against *p*'s superassertibility is an argument against *p*'s being currently justified. So error theory must show that all or most assertions actually made are not justified by the standards of the discourse. This might be because we don't realize that the standards are not coherent, so there is in fact no such thing as being justified according to those standards. Alternatively, we might be under a misapprehension about having met those standards.

Paradigmatic error theorists such as Mackie and Field assume that we do in fact satisfy the standards of warrant acknowledged as governing a discourse. Hence they must presuppose that truth is not superassertibility and their error theories must include a case for this presupposition. How does one provide such a case? There is one apparently relatively unproblematic type of argument, applying to fiction. On Wright's view, discourse "both within and about a fiction characteristically" has the syntactic features of minimal assertion, "and is subject to a high degree of internal discipline: there are many claims about Hamlet which are determinately correct, and many which are determinately incorrect" (2003: 63).³⁰ So fiction contains minimally truth-apt assertions. But we don't, by and large, take sentences occurring in fiction or (some) sentences we utter about fiction to be (literally) true: for instance, we don't think that there existed two people, Charles

²⁹ See Jackson et al. (1994).

³⁰ It's not clear to me in what sense, if any, the production of the sentences of a novel or poem by its author is governed by acknowledged standards of correctness.

and Emilia Gould, who lived in a Latin American country called “Costaguana,” so we don’t think that either Conrad’s sentence, “Mrs. Gould loved the patio of her Spanish house,” or our sentence about *Nostromo*, “Charles Gould met his wife in Italy” is literally true.³¹ But this form of argument is of no use to error theory, because no illusion is involved. The situation is not that we wrongly believe many sentences of fiction to be literally true, such that metaphysical inquiry is needed to remove the scales from our eyes.

It follows that an error-theoretic case for distinguishing truth from superassertibility must proceed under the constraint that we don’t already take most sentences of the discourse to be false. But recall that we’re proceeding under the assumption that our (purported) assertions *are* warranted by the standards governing the discourse. So there’s no reason to think that our assertions are not superassertible, and hence no reason to think that we’re mostly wrong in taking our assertions to be true, *if* truth is superassertibility. That is to say, it’s possible to interpret truth for the discourse in such a way that we are not mostly mistaken. The principle of charity, then, enjoins that we take the burden of proof to be on error-theories, to show that our discourse has ontological presuppositions independent of the standard we take to govern assertions, and that they are not fulfilled.

I turn now to discuss two attempts to shoulder this burden. First, Frank Jackson produces two types of examples of “discourses that have coherent, articulated and acknowledged standards of proper use, and yet many or most of the statements that satisfy those standards . . . are false” (1994: 167). One type consists of false scientific theories such as Newtonian physics and Ptolemaic cosmology. The other type are models, known to be false, used in physical explanations, such as ideal gas theory. Now, does the first type of example demonstrate the possibility of the failure of discourse-independent ontological presuppositions for the statements of, say, Newtonian mechanics? It’s not clear that it does, because we came to recognize the falsity of Newtonian physics by doing physics, in other words on the basis of standards of warrant governing the discourse of physics. Moreover, for this reason, what we have recognized is also that these statements are not superassertible. The second type of example is in one respect much more complex, since the philosophical status of models in science is

³¹ Miller (2001) argues that since truth of fiction cannot be evidence transcendent, superassertibility has to be a model of truth for fictional discourse. But it’s not clear that being able to model truth for a discourse is sufficient to be its norm of truth, when practitioners of that discourse would acknowledge that reason to believe a sentence to be superassertible is consistent with its literal falsity.

Fictional discourse is the basis of an objection to the entire minimalist framework advanced by Divers and Miller (1995). Their argument is that since fictional discourse is minimally assertoric, assertoric minimalism implies that it consists largely of true statements, and so commits us, implausibly, to an ontology containing fictional characters. Wright’s account of truth for fiction, together with the point made in the last paragraph suffices to block this consequence.

In any event, Wringe (2008) proposes that one could accept this ontology but retain or rescue plausibility by arguing that the notion of truth applying to fiction fails to be objective by the criterion of cognitive command.

controversial.³² But in another respect it is clear that these examples no more support error theories than does the case of fictional discourse: the models are known to be false.

Second, Timothy Williamson (1994) presents a case in which, to use Frege's words, we have unwittingly lapsed into fiction (1984: 367): speakers take "Eldorado is crowded" to be a genuine statement and are wrong because "Eldorado" fails to refer. In this case it is fairly clear what the content of the illusion is: the speakers believe wrongly that the expression "Eldorado" is a name of a city. So it is clear enough what kinds of investigation would force them to conclude that this belief is false. But do we have, similarly, a clear idea of what we believe, for example, the numeral "1" refers to? Or what we believe "is evil" expresses? So it is not at all clear what our ontological presuppositions *are*, for our utterances to be genuine assertions, much less that we have reason to believe that they are not satisfied.

There are two main strands to Wright's response to error theories. First, pervasive metaphysical illusion seem *prima facie* less plausible than taking our ordinary conceptions of the truth-aptness of what we say at face value. Second, whenever a ground for suspecting massive metaphysical error is produced, it is shown to be interpretable as a possibility internal to our assertoric practices, and so we need not take the path of error theory. So there is, obviously, no refutation of the ontological approach, only repeated attempts to show that it, and its costs, are not compulsory.

16.7 A CONCLUDING HISTORICAL REMARK

There is much in the semantic approach that I have not discussed. I have not touched on semantic metaphysical arguments and conclusions about various particular aspects of reality: Dummett and Wright on the past, Dummett on causation, Wright on morality, necessity, and psychology.³³ Nor have I examined the details of Dummett's views of theories of meaning or Wright's treatment of theories of truth.³⁴ Perhaps most disappointingly, I've said little in evaluation of the semantic arguments. But my goal is to outline a defensible and more or less unified framework for the semantic approach, one which removes some misconceptions, and sharpens and clarifies the philosophical motivations of Dummett's and Wright's work.

I conclude by making explicit what the reader has perhaps suspected all along: the semantic approach has affinities with the views of Kant and Carnap. Kant's transcendental idealism insists that metaphysical theorizing about reality has to proceed through an account of the conditions under which it is possible for us to have experience or knowledge of that reality. Similarly, the semantic approach proposes that metaphysical

³² For a good survey, see Frigg and Hartmann (2006).

³³ See Dummett (1978; 1991); Wright (1980; 1993; 2003).

³⁴ See Dummett (1981b; 1991: chs. 1–10); Wright (1992; 2003: Part IV).

theorizing about reality proceed through an account of the conditions under which we represent reality in thought or discourse. The Carnapian element in the semantic approach lies in the fact that the semantic method of metaphysics is a proposal. Kant surely claims to have an argument for transcendental idealism as the only correct path in metaphysics. The semantic method, in contrast, suggests explications, in Carnap's (1950) sense, of metaphysical disputes, which (perhaps) have the advantage of being subject to clearer standards of correctness, and which neither grossly distort what traditional metaphysicians take themselves to be arguing about nor miss out on any significant aspects of ontology.

REFERENCES

- Blackburn, S. (1989). "Manifesting Realism," *Midwest Studies in Philosophy* 14: 29–47.
- Blackburn, S. (1998). "Realism and Truth: Wittgenstein, Wright, Rorty and Minimalism," *Mind* 107: 157–81.
- Brouwer, L. E. J. (1975) [1908]. "The Unreliability of the Logical Principles," English translation of "De onbetrouwbaarheid der logische principes," in A. Heyting (ed.), *L. E. J. Brouwer: Collected Works vol. 1: Philosophy and Foundations of Mathematics*. Amsterdam: Elsevier, 107–11.
- Byrne, D. (2005). "Compositionality and the Manifestation Challenge," *Synthese* 144: 101–36.
- Carnap, R. (1950). *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- Devitt, M. (1983). "Dummett's Anti-Realism," *Journal of Philosophy* 80: 73–99.
- Devitt, M. (1991). *Realism and Truth*, 2nd edn. Oxford: Blackwell.
- Devitt, M. and K. Sterelny. (1987). *Language and Reality: An Introduction to the Philosophy of Language*. Cambridge, MA: MIT Press.
- Divers, J. and A. Miller. (1995). "Minimalism and the Unbearable Lightness of Being," *Philosophical Papers* 24: 127–39.
- Dummett, M. A. E. (1963). "Realism," in Dummett (1978), 145–65.
- Dummett, M. A. E. (1973). "The Philosophical Basis of Intuitionistic Logic," in Dummett (1978), 215–47.
- Dummett, M. A. E. (1977). *Elements of Intuitionism*. Oxford: Oxford University Press.
- Dummett, M. A. E. (1978). *Truth and other Enigmas*. London: Duckworth.
- Dummett, M. A. E. (1981a). *The Interpretation of Frege's Philosophy*. London: Duckworth.
- Dummett, M. A. E. (1981b). *Frege: Philosophy of Language*, 2nd edn. London: Duckworth.
- Dummett, M. A. E. (1982). "Realism," in M. A. E. Dummett, *The Seas of Language*. Oxford: Oxford University Press, 230–76.
- Dummett, M. A. E. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Dummett, M. A. E. (1993). "Realism and Anti-Realism," in Dummett, *The Seas of Language*. Oxford: Oxford University Press, 462–77.
- Field, H. H. (1980). *Science without Numbers: A Defence of Nominalism*. Princeton: Princeton University Press.
- Frege, G. (1984). *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Blackwell.

- Frigg, R. and S. Hartmann. (2006). "Models in Science," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2006 Edition), <http://plato.stanford.edu/entries/models-science/>
- Hale, B. (1997). "Realism and its Oppositions," in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell, 271–308.
- Hale, B. and C. Wright. (2001). *The Reason's Proper Study: Essays towards a neo-Fregean Philosophy of Mathematics*. Oxford: Oxford University Press.
- Hand, M. (1998). "Radical Antirealism and Wright's Antideflationary Argument," *Noûs* 32: 320–30.
- Horgan, T. (1996). "The Perils of Epistemic Reductionism," *Philosophy and Phenomenological Research* 56: 891–7.
- Horwich, P. (1996). "Realism minus Truth," *Philosophy and Phenomenological Research* 56: 877–83.
- Jackson, F. (1994). "Review of C. Wright, *Truth and Objectivity*," *Philosophical Books* 35: 162–9.
- Jackson, F., G. Oppy, and M. Smith (1994). "Minimalism and Truth-Aptness," *Mind* 103: 287–302.
- Kirkham, R. I. (1989). "What Dummett Says about Truth and Linguistic Competence," *Mind* 98: 207–24.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. New York: Penguin.
- McDowell, J. (1981). "Anti-Realism and the Epistemology of Understanding," in J. McDowell, *Meaning, Knowledge and Reality*, Cambridge, MA: Harvard University Press, 314–43.
- McGinn, C. (1980). "Truth and Use," in M. Platts (ed.), *Reference, Truth and Reality*. London: Routledge, 19–40.
- Miller, A. (2001). "On Wright's Argument against Deflationism," *Philosophical Quarterly* 51: 527–31.
- Miller, A. (2002a). "Wright's Argument against Error-Theories," *Analysis* 62: 98–103.
- Miller, A. (2002b). "What Is the Manifestation Argument?," *Pacific Philosophical Quarterly* 83: 352–83.
- Miller, A. (2005). "Realism," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2005 Edition), <http://plato.stanford.edu/entries/realism/>
- Miller, A. (2006). "Realism and Antirealism," in E. Lepore and B. Smith (eds.), *Handbook of Philosophy of Language*. Oxford: Oxford University Press, 983–1005.
- Pettit, P. (1996). "Realism and Truth: A Comment on Crispin Wright's *Truth and Objectivity*," *Philosophy and Phenomenological Research* 56: 883–90.
- Putnam, H. (1983). *Realism and Reason*. Cambridge: Cambridge University Press.
- Putnam, H. (2001). "When Evidence Transcendence is not Malign: A Reply to Crispin Wright," *Journal of Philosophy* 98: 594–600.
- Rosen, G. (1994). "Objectivity and Modern Idealism: What is the Question?," in M. Michael and J. O'Leary-Hawthorne (eds.), *Philosophy in Mind: The Place of Philosophy in the Study of Mind*. Dordrecht: Kluwer, 277–319.
- Rosen, G. (1995). "The Shoals of Language: Michael Dummett's *The Seas of Language*," *Mind* 104: 599–609.
- Sainsbury, M. (1996). "Crispin Wright: *Truth and Objectivity*," *Philosophy and Phenomenological Research* 56: 899–904.
- Shapiro, S. and W. W. Taschek. (1996). "Intuitionism, Pluralism, and Cognitive Command," *Journal of Philosophy* 93: 74–88.

- Shieh, S. (1998a). "Undecidability in Anti-Realism," *Philosophia Mathematica* 6: 324–33.
- Shieh, S. (1998b). "On the Conceptual Foundations of Anti-Realism," *Synthese* 115: 33–70.
- Shieh, S. (1999). "What Anti-Realist Intuitionism Could Not Be," *Pacific Philosophical Quarterly* 80: 77–102.
- Shieh, S. (2012). "A Tale of Two Anti-Realisms," in S. Rahman, G. Primiero, and M. Marion (eds.), *The Realism–Antirealism Debate in the Age of Alternative Logics*. Dordrecht: Springer, 285–99.
- Shieh, S. (forthcoming). "Wright and Revisionism," in A. Miller and A. Coliva (eds.), *Essays for Crispin Wright: Logic, Language and Mathematics*. Oxford: Oxford University Press.
- Stalnaker, R. (1997). "Reference and Necessity," in B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*. Oxford: Blackwell, 534–54.
- Strawson, P. F. (1977). "Scruton and Wright on Anti-Realism etc.," *Proceedings of the Aristotelian Society* 77: 15–21.
- Tarski, A. (1954–5). "Contributions to the Theory of Models, I, II, III," *Indagationes Mathematicae* 16: 572–88; 17: 56–64.
- Taylor, B. (1987). "The Truth in Realism," *Revue internationale de philosophie* 41: 45–63.
- Tennant, N. (1995). "On Negation, Truth, and Warranted Assertibility," *Analysis* 55: 98–104.
- Thomason, R. H. (1970). "Indeterminist Time and Truth-Value Gaps," *Theoria* 36: 264–81.
- Van Cleve, J. (1996). "Minimal Truth is Realist Truth," *Philosophy and Phenomenological Research* 56: 869–75.
- Williamson, T. (1994). "Crispin Wright's Truth and Objectivity," *International Journal of Philosophical Studies* 2: 130–44.
- Williamson, T. (1996). "Unreflective Realism," *Philosophy and Phenomenological Research*, 56: 905–9.
- Wright, C. (1980). *Wittgenstein on the Foundations of Mathematics*. Cambridge, MA: Harvard University Press.
- Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Wright, C. (1993). *Realism, Meaning, and Truth*, 2nd revised edn. Oxford: Blackwell.
- Wright, C. (2001). "On Being in a Quandary: Relativism Vagueness Logical Revisionism," *Mind* 110: 45–98.
- Wright, C. (2003). *Saving the Differences: Essays on Themes from Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Wringe, B. (2008). "Making the Lightness of Being Bearable: Arithmetical Platonism, Fictional Realism and Cognitive Command," *Canadian Journal of Philosophy* 38: 453–88.

CHAPTER 17

DEFLATIONIST TRUTH

JODY AZZOUNI

17.1 THE TRUTH SCHEMA AND REDUNDANCY VIEWS OF TRUTH

CONSIDER the following statements:

(QUOTE PLUS)	" $2 + 2 = 4$," is true.
(PLUS)	That $2 + 2 = 4$ is true. ¹
(QUOTE ALBERT EINSTEIN)	"Albert Einstein was born in New Jersey," is false.
(ALBERT EINSTEIN)	That Albert Einstein was born in New Jersey is false.

The grammatical form of such statements makes it natural to think that "true" and "false" are properties of sentences—or something like sentences (propositions, thoughts, etc., see e.g. Künne 2003: ch. 5); "something like sentences," both because there can be debate as to the references the items "is true" and "is false" are appended to, and because locutions like

The proposition that Albert Einstein was born in New Jersey is false

seem to expressly attribute truth and falsity to "propositions." I'll continue to use the neutral term, "statement," leaving aside the question of what, exactly, the "vehicles" of truth and falsity are. We can still ask, "What kind of properties are truth and falsity?" Consider the following:

(QUOTE TRUISM)	"Snow is white," is true if and only if snow is white.
(TRUISM)	That snow is white is true if and only if snow is white.

¹ This can read awkwardly. It's entirely natural-sounding, however, if one stresses "that," and pauses ever so slightly after "4" before stressing (slightly) "is true."

These are two versions of only one of the many instances of the “T-schema” in English. Each such instance is generated by substituting some other English sentence for “snow is white” everywhere in QUOTE TRUISM and everywhere in TRUISM. Thus,

(QUOTE WOLVES)	“Wolves are smart,” is true if and only if wolves are smart.
(WOLVES)	That wolves are smart is true if and only if wolves are smart.

In the rest of this chapter, I’ll largely focus on the nonquote-versions of the T-schema rather than on the quote-versions. The motivation for this apparently notational choice will emerge shortly.²

Some philosophers use the instances of the T-schema to argue that the truth-property is a correspondence property. There are many theories, however, about what truth-as-correspondence comes to. Some think that property can be read off from the instances of the T-schema. TRUISM, for example, indicates an identification of the truth-property attributed to the relatum of “that snow is white,” on the left, and the fact expressed by “snow is white,” on the right. Some philosophers introduce a “state of affairs” as the right-hand relatum of “snow is white,” or an (ordered) collection of objects and properties as that right-hand relatum.

Other answers to the question of what property truth is are found in the tradition: “coherence” and “utility,” for example. Given the aims of this chapter, we won’t probe these answers further.³ What’s important is the *kind* of answers these are. True statements are taken to share some property *P* that can be characterized independently of the concept of truth. A description of *P*, therefore, provides necessary and sufficient conditions on a statement being true—necessary and sufficient conditions that don’t (even implicitly) involve the concept of truth. I refer to such a description of *P* as a *substantialist* theory of truths. Notice that substantialist theories of truths are global ones, about all truths (e.g. that all truths correspond to facts: Russell 1912). Opposed to these are *pluralistic* theories of truths—ones that divide the sets of truths into different groups of discourse where the truths in any one such group share a truth-property that’s different from the truth-properties shared by truths in other groups of discourse (see e.g. Lynch 2004). Finally, there is the *deflationist* theory of truths. This denies that there can be any useful categorizations of truths in terms of shared substantial truth-properties. These different theories of truths will be discussed in section 17.4. In the meantime, I take up other strands in the deflationist tradition in order (eventually) to show how they are connected to issues regarding truth-properties and regarding substantialist and deflationist theories of truths.

Let’s return to the evidence from ordinary usage that started this line of thought about truth-properties. The deflationist tradition draws strikingly different conclusions from that evidence. Consider PLUS again.

(PLUS)	That $2 + 2 = 4$ is true.
--------	---------------------------

² See fn. 7.

³ However see section 17.4, especially fn. 20.

Many philosophers over the course of the last century have noted that speakers don't have to utter PLUS; they can utter the truncated and more convenient

$$2 + 2 = 4$$

instead. With other assumptions, this seems to yield (e.g. Ramsey 1927):

(REDUNDANCY) The statement that a statement S is true comes to nothing more than S itself.

REDUNDANCY is open to a number of interpretations that I'll eventually coalesce into two distinct positions. To begin with, the following claims have been made by one or another deflationistically-inclined philosopher. (1) The instances of the T-schema constitute the full meaning of "true"—together they provide everything needed to understand "true"; (2) a sincere normal assertion of a statement undertakes a commitment to the truth of that statement (and vice versa); (3) to say that a statement is true is to say no more than the statement itself; (4) "true" is a redundant locution the content of which is fully contained within our utterance-practices (and their norms) with respect to statements in which "true" doesn't appear.

These glosses on REDUNDANCY differ in degrees of philosophical *rashness*. (4)—and (3) for that matter—it may be claimed, are refuted by

(BLIND₁) Everything Aristotle wrote about lobsters is false.

(BLIND₂) Some of what Claire said yesterday is true.

The instances of the T-schema, *prima facie*, can't be used to eliminate the uses of "true" and "false" from BLIND₁ and BLIND₂, and so "true" and "false," *prima facie*, can't be described as "redundant locutions." No more does it seem that in saying either BLIND₁ or BLIND₂ one is actually *saying* the statements or negations of (or conjunctions/disjunctions of) what Aristotle wrote about lobsters, or what Claire said yesterday. Such interpretations of BLIND₁ and BLIND₂ seem to misconstrue the evident meanings of these statements. Rather, these locutions—*blind truth and falsity ascriptions*⁴—respectively say exactly what they appear to say, that each statement from among the statements described as what "Aristotle wrote about lobsters" is false, and that at least one statement from among the statements described as "what Claire said yesterday" is true.

⁴ For many years I've used "blind" to describe such locutions. Some philosophers now describe them as "opaque," as in "opaque truth ascriptions." I chose "blind" instead of "opaque" deliberately to avoid terminological conflict with "opaque" as already widely used in philosophy of language—e.g. as a characterization of belief contexts.

17.2 T-SCHEMA AND SEMANTIC DESCENT DEFLATIONISM

Philosophers intent on a deflationist reading of BLIND₁ and BLIND₂ aren't without resources for a response. It's been noticed that uses of "true" and "false," when appended to quantifiers (ranging over statements) that are coupled to descriptions of such statements, can be captured by a formalism that allows infinitely long statements (see e.g. Quine 1970; Leeds 1978; Putnam 1978; Field 2001). Let⁵ $S_1, S_2, S_3, \dots, S_n, \dots$ be a list of (all) statements, and suppose we presume a canonical list of sentences of English: $s_1, s_2, s_3, \dots, s_n, \dots$, such that for $i = 1, 2, \dots, n, \dots$, each S_i is expressed by s_i .⁶ Instead of

(BLIND₁) Everything Aristotle wrote about lobsters is false,

one now says (but only in the fullness of time),

(T-INFINITE₁) (If Aristotle wrote about lobsters that s_1 , then it is false that s_1) & (if Aristotle wrote about lobsters that s_2 , then it is false that s_2) ... & ...

Applying an appropriate instance of the T-schema to each clause of T-INFINITE₁ gives us (in the fullness of time),

(INFINITE₁) (If Aristotle wrote about lobsters that s_1 , then $\neg s_1$) & (if Aristotle wrote about lobsters that s_2 , then $\neg s_2$) ... & ...⁷

⁵ My use of the variables: S_1, S_2, S_3, \dots , etc., s_1, s_2, s_3, \dots , etc., introduced to stand respectively for statements and sentences, implicitly operates thus: a quotation mark followed by such a letter followed by another quotation mark: —" S_1 "—, isn't a name of the subscripted capitalized nineteenth letter of the alphabet, but a name of the statement that this item stands for. So too: —" $(S_1 \& S_2)$ "—, is the name of the item: left parenthesis concatenated with the statement the capitalized nineteenth letter subscripted with the numeral "1" stands for concatenated with the ampersand concatenated with the statement the capitalized nineteenth letter subscripted with the numeral "2" stands for concatenated with the right parenthesis. Similar remarks hold of the variables " A ," " B ," etc., in section 17.3, that stand in general for statements. I use quote marks to indicate conventionally and informally the presence of semantic ascent; see the discussion of semantic ascent and descent that follows.

⁶ This is a major assumption, and it may be ultimately unsustainable—no matter how it is tinkered with. (I actually think it *is* unsustainable. This is one motivation I have for preferring the to-be-described "semantic-descent" deflationism to the currently-under-discussion "T-schema deflationism.") There are issues with context-dependent and singular sentences; there are issues about whether it can be presumed that there are available sentences in the speaker's language that capture through translation the meanings of sentences in other languages; finally, there are issues about how this approach, or any in its neighborhood, can handle the family of "self-referential" statements. I won't discuss this constellation of interconnected challenges now. See Azzouni (2001; 2006) for some discussion of them.

⁷ As with PLUS this may seem to read unnaturally. Imagine it this way: "If Aristotle wrote that lobsters are red about lobsters then it is not the case that lobsters are red; and if ... " Notice that if we approached BLIND₁ via the quote-version of the T-schema, we would have to formulate INFINITE₁ in the form: "If Aristotle wrote about lobsters ' s_1 ', then $\neg s_1$) & ...," which wrongly attributes the writing of English sentences to Aristotle.

Instead of BLIND₂, we can similarly write using a corresponding T-INFINITE₂,

(INFINITE₂) (Claire said that s_1 yesterday & s_1) or (Claire said that s_2 yesterday & s_2)
 ... or

INFINITE₁ and INFINITE₂ can, in turn, be captured by *finitely-long* locutions via the substitutional quantifiers (x) , $(\exists x)$, where the substituents are the statements of English. Then we have

(SUBST₁) (x) (If Aristotle wrote that x about lobsters, then $\neg x$),
 (SUBST₂) $(\exists x)$ (Claire said that x yesterday & x).

If we systematically replace blind truth and falsity ascriptions with statements like INFINITE₁ and INFINITE₂, or with statements like SUBST₁ and SUBST₂, it seems we have shown that “true” and “false” are redundant. Any natural language could have dispensed with such idioms if the language had allowed infinitely-long conjunctions (and disjunctions) or substitutional quantifiers with natural-language sentences as substituents. We can also take the foregoing as an explanation of how we understand blind truth (and falsity) ascriptions (i) given that an understanding of “true” amounts to the instances of the T-schema and (ii) given that an understanding of the quantifiers amounts to infinite disjunctions and conjunctions.

It may be justly objected that if English *doesn't* allow infinitely-long disjunctions and conjunctions and if it *doesn't* allow substitutional quantifiers, then it can hardly be claimed that “true” and “false” are redundant. Maybe so, but deflationists can retort that it's been made clear by SUBST₁, SUBST₂, INFINITE₁, and INFINITE₂, that the concepts involved in blind truth and falsity ascriptions don't—strictly speaking—go beyond the instances of the T-schema when some method of handling reference to indeterminate *numbers* of statements is supplied. A careful statement of the deflationist claim should replace a rash description of the eliminability of “true” and “false” with a more nuanced characterization of the slender conceptual resources needed for the successful understanding of the natural-language role of “true” and “false.”

It's appropriate to distinguish between *two* possible deflationist views: *T-schema deflationism* and *semantic-descent deflationism*. The T-schema deflationist claims that

(TRUISM) That snow is white is true if and only if snow is white,

and its T-schema brethren (generated by substituting each s_i from the canonical list of sentences for “Snow is white” in TRUISM) constitute partial definitions of “is true” *in the strict sense* that we understand *each use* of “true” in sentences of the form

(PLUS) That $2 + 2 = 4$ is true,
 (ALBERT EINSTEIN) That Albert Einstein was born in New Jersey is false,

only because we have an appropriate instance of the T-schema that (coupled with negation in the second case) we can use to eliminate “is true” (or “is false”). We understand “That $2 + 2 = 4$ is true” *to mean* “ $2 + 2 = 4$.” We understand “That Albert Einstein was born in New Jersey is false” *to mean* “It is not the case that Albert Einstein was born in New Jersey.” We, therefore, understand

(BLIND₁) Everything Aristotle wrote about lobsters is false,

(BLIND₂) Some of what Claire said yesterday is true,

only because we recognize that they are (respectively) abbreviations of T-INFINITE₁ and T-INFINITE₂, and therefore we recognize the uses of “is true” in BLIND₁ and BLIND₂ to be uses to which instances of the T-schema apply—despite the contrary appearances induced by the abbreviations. Alternatively, we understand BLIND₁ and BLIND₂ because the truth conditions for these statements are given in terms of substitutions from the canonical sentences of English, and we understand what such substitutions entail only because of our understanding of the instances of the T-schema. If, however, BLIND₁ and BLIND₂ are taken to be ordinary quantifiers, quantifiers like

Everything Aristotle did to lobsters was extremely cruel,

Some of what Claire ate yesterday made her ill,

that is, then the T-schema deflationist *can't explain* how anyone is to understand BLIND₁ and BLIND₂. His way of doing so *requires* reconstruing the quantifiers used with the word “true” either as abbreviations of the form T-INFINITE₁ and T-INFINITE₂, or as substitutional quantifiers with the canonical list of English sentences as substituents.

Many T-schema deflationists claim that “true” is a device of generalization (e.g. Horwich 1998). It cannot be so construed, even given the assumptions of such deflationists: on their view, it's a word elucidated on a case-by-case basis (via each instance of the T-schema) by an identification of the statement expressed by the sentence—when sandwiched between “that” on the one side, and “is true” on the other—with the statement expressed by that sentence. Hardly a device of generalization! But perhaps the fact that T-schema deflationism requires the reconstruction of the quantifiers occurring in blind truth and falsity ascriptions invites the “device of generalization” mischaracterization of the word “true.”⁸

⁸ Another explanation for this misconstrual of “true” may be that if one has certain generalizing devices—e.g. resources for stating infinite disjunctions and conjunctions—then “true” isn't needed: one can handle blind truth ascriptions directly, as INFINITE₁ and INFINITE₂ do. One can then see the reintroduction of “true” (in BLIND₁ and BLIND₂) as the introduction of an abbreviation-device for these disjunctions and conjunctions (see e.g. Field 2001). But this too misreads what “true” is doing in BLIND₁ and BLIND₂: it's the *quantifiers* in those statements that have the generalization role, while the defined “true” functions purely in a semantic descent role.

It should be noted that what's required of T-schema deflationism is that *all* the generalized quantifiers are required to be reconstrued in infinitary fashion. Consider:

(MOST) Most of what Claire said yesterday is true.

This can be managed with an infinitary statement, i.e. a disjunction of the form:

(Claire said yesterday that s_1 & said s_2 & said s_3 & s_1 & s_2) or (Claire said yesterday that s_1 & said s_2 & said s_3 & s_1 & s_3) or ...

coupled with an infinite conjunction of the form:

\neg (Claire said yesterday that s_1 & $\neg s_1$ & Claire said yesterday that s_1 and said s_2 & $\neg s_1$...).

Similar maneuvers are needed for other quantifier expressions such as, "Exactly three," "Four times as many," "Roughly half," and so on.⁹

The semantic-descent deflationist, by contrast, urges a deflationism that doesn't require melding the understanding of the truth idiom in blind truth ascriptions case-by-case with instances of the T-schema and the evaporation of all the quantifiers into infinite statements of various sorts. She, correspondingly, denies that the instances of the T-schema are partial definitions of "true." She takes "true" to be a device of semantic descent: it correlates statements with nominalizations—descriptions or names—of those statements, so that statements containing the latter descriptions or names *can stand stead for the statements themselves*. This, she notes, is *stated* by the instances of the quote-T-schema *only with* respect to quote names, and with the T-schema *only with* respect to nominalizations of the form: that s . That the same is true of other names of such statements or of descriptions of (collections of) statements is because of co-referring substitution conventions for names, and because of similar conventions for descriptions. The instances of the T-schema, therefore, indicate how "true" functions as a device of semantic descent for a certain class of statement: they fix the statements that "that"-phrases (and quote-names) appended with "is true" are to stand *stead for*.

Apart from denying that instances of the T-schema are abbreviatory partial definitions of "true," asserting rather that they *illustrate* "true" as a device of semantic-descent, the semantic-descent deflationist officially holds that devices of generalization are *quantifiers* (although infinite sentences can sometimes do the same job). Some

⁹ My thanks to Dan Greco for pointing out the disjunction strategy for "most." Although he didn't mention the necessary closure conjunction clause, I assume it's part of his strategy. Notice, though, that if we say "Most of what God said yesterday is true," and we allow God the capacity to utter infinitely many statements in a day, the cardinality of the disjuncts (and conjuncts) goes unpleasantly continuum-many. One advantage of *quantifiers* is their ability to represent cases without having to code cardinality explicitly beyond how it arises explicitly in the quantifiers themselves.

devices of generalization obviate the need for a device of semantic descent—a truth predicate. This is true if a language has the resources to state INFINITE₁ and INFINITE₂, or SUBST₁ and SUBST₂. Semantic descent is explicitly handled by INFINITE₁ and INFINITE₂ via the canonical *list*—in conjunctions and disjunctions—of those statements mentioned and then used; semantic descent is handled by SUBST₁ and SUBST₂ by substitutional quantification within and without “that”-phrases. Even better ways of incorporating semantic descent into devices of quantification are possible—by allowing quantifiers to govern variables occurring *simultaneously* in sentential, and in singular term, positions.¹⁰

If a device of generalization itself handles semantic descent, then “true” and “false” aren’t needed. “True” and “false” are essential to natural languages at least because natural-language quantification over statements does *not* involve quantification devices that themselves handle semantic descent: quantification in natural languages is coupled with descriptions that mention such statements, but don’t use them. The truth predicate in natural languages, therefore, is a *predicate*—a genuine predicate—that enables us to describe these statements in a way that can stand stead for (in the sense indicated by the instances of the T-schema) using them.

Most truth deflationists—and importantly, most critics of truth deflationism—claim that truth deflationism *requires* of the quantification devices, used to replace the combination in natural languages of the truth predicate and ordinary quantifiers, that they must be devices of infinite disjunction and conjunction, or must be abbreviations for such. This misapprehension arises from those philosophers either overlooking the possibility of semantic-descent deflationism, or failing to distinguish it clearly from T-schema deflationism (e.g. as in many of the articles in Blackburn and Simons 1999, or Schantz 2001).

One reason to prefer semantic-descent deflationism over T-schema deflationism is that it’s the *truth predicate* that’s the target of the philosophical analysis engaged in by truth deflationists (generally), and not the accompanying quantifier that ranges—however it does so—over statements. It’s a serious liability of *any* position to require an ad hoc reconstrual of the quantifiers that occur in blind truth and falsity ascriptions. For—at least as far as ordinary languages are concerned—such quantifiers appear to be exactly the same ones as appear elsewhere in English, and to belong to a more general family of quantifiers that are similar in their semantics. This places a heavy empirical demand on the T-schema deflationist.

A related obstacle facing T-schema deflationism is that such a view describes the instances of the T-schema as indicating their left sides simply amount to their right sides. But this obscures the evident resemblance between LEFT TRUISM

(LEFT TRUISM) That snow is white is true,

¹⁰ See the discussion of anaphorically unrestricted quantifiers in Azzouni (2001; 2006). One reason anaphorically unrestricted quantifiers are better than substitutional quantifiers is that they dispense with the need for a canonical list of sentences to function as substituents. Recall fn. 5.

and

(HARD) That snow is white is hard to believe.

HARD and LEFT TRUISM are about the same item, although they make different claims about it. “Snow is white,” on the other hand, doesn’t talk about that item at all.

It’s natural to ask, are there philosophical advantages to T-schema deflationism, that the semantic-descent deflationist is giving up? One apparent advantage is that T-schema deflationism tells a particularly simple story about the understanding of “true.” One’s understanding stems solely from the instances of the T-schema. The semantic-descent deflationist instead reads our understanding of the truth predicate directly from its semantic descent role. She notes that Tarski’s definition of truth, for example, directly utilizes descriptions of the sentences of a language to axiomatize (or, in some cases, to define) the needed descent device. The T-schema instances then emerge naturally as *corollaries* on Tarski’s approach. In cases where the generalizing devices themselves are such that a truth predicate can be defined from them,¹¹ an understanding of the truth predicate follows from an antecedent understanding of those generalizing devices. Correlatively, recognition of the truth of the instances of the T-schema follows from an antecedent understanding of these generalizing devices rather than that understanding of such generalizing devices (as they are used in blind truth ascriptions) following from an understanding of the instances of the T-schema.¹²

Some truth deflationists—and many critics of truth deflationism—claim that the deflationist view of truth is (must be) that “true” is a “device of endorsement”: we have such a term because of a need to endorse propositions indirectly. Although “true” is often *used* to indirectly endorse other statements, it’s no more (and no less) a “device” of endorsement than *any* statement is a “device” of endorsement for its own content. Characterizing a truth predicate as a device of endorsement is failing to recognize that the way a piece of language is often—or even typically—*used* needn’t be what it *is*. (That

¹¹ E.g. if a substitutional quantifier, as above, is available, one can axiomatize a truth predicate like so: $(x)(\text{True} \text{that} x \text{ iff } x)$, where “*that*” is a nominal expression bound by the quantifier. A definition of the truth predicate, therefore, is easily within reach.

¹² T-schema deflationism gives a “bottom-up” characterization of the understanding of truth: one starts with a grasp of the instances of the T-schema and tries to inflate that into a full understanding of the role of the truth predicate in every sentential context. Semantic-descent deflationism employs a “top-down” characterization of understanding truth: one starts with a semantic descent grasp of truth *in tandem* with one’s grasp of one or another generalizing device, and one’s grasp of the instances of the T-schema follows as a corollary.

This point is worth adding: quantifiers and logical connectives can’t *all* be defined. Some set or another of them must be taken as primitive. Their being taken as primitive is manifested by a characterization of one’s understanding of them in formulations that utilize them (e.g. in truth-conditional meta-language statements). This is the case with “true” as well. It may be taken either as a primitive notion (and therefore to be axiomatized), or it may be defined from other notions—quantifiers of some sort.

“true” and “false” are devices *for* blind truth and falsity ascriptions doesn’t make them devices *of* blind truth and falsity ascription.) Were a truth predicate *just* a device of endorsement, it would be impossible to use “true” in pretence, as when an actor says, “It’s true that there are unicorns.” By saying this, the actor *would* endorse the claim that there are unicorns. So too, it would be impossible to use the statement

(PLUS) That $2 + 2 = 4$ is true,

as an *example* (as I’ve just done). For in doing so, the statement “ $2 + 2 = 4$ ” would be *automatically* endorsed, and would thus fail to be a *mere* example.

The judicious deflationist claim (again) should be this: “true” and “false” function as devices of semantic descent. When coupled with quantifiers and descriptions, they can, of course, be *used* to endorse (or repudiate) statements indirectly. The judicious deflationist could claim that the *reason* why “true” and “false” occur in natural languages is to enable the expression of blind truth and falsity ascriptions such as BLIND₁ and BLIND₂. But it’s a mistake to *add* that “true” and “false” *are* devices of blind truth and falsity ascription. They are not. It’s their role as devices of semantic descent that enables them to be used (with other resources) to express blind truth and falsity ascriptions.

Call the foregoing description of the semantic descent role of “true,” and “false,” and how it can be extended to blind truth ascription, “the semantic-descent deflationist theory of truth.” It’s a theory of a piece of ordinary language—“true,” and of the other idioms that can stand in for “true” in ordinary language. So perhaps it’s better described as “the semantic-descent deflationist theory of ‘true,’” and I’ll do so from now on. This theory of “true” explains one thing that this particular piece of language is needed for, and what rules govern our understanding of this role that it has. The resulting theory, therefore, can be quite weak. We know for instance that the moon is not true; we also know, presumably, that the conjunction of two statements is true if and only if both conjuncts are true; the semantic-descent deflationist theory of “true” needn’t be strong enough to prove either. As far as *it* is concerned, these claims could be false. That’s not a problem. “True” here is a device of semantic descent, and even when that characterization of it is further coupled with the characterization of quantifier, the result won’t necessarily give us details about *truth vehicles* (what the quantifiers in fact range over), or other details of the logic that both devices are used with (e.g. whether it’s classical, intuitionistic, or whatnot, *contra* Gupta 1993; Ketland 1999). All of this is to be supplied by supplementary theories that only in tandem with the characterization of “true” and the quantifiers may yield results such as that the conjunction of two statements is true if and only if both conjuncts are true (if that, in fact, *is* true), or similarly, that the moon isn’t true (if, in fact, the moon isn’t even a sentence). It’s *also* compatible with the claim that the natural-language idiom “true” has other roles—as in a “true” friend.

The semantic-descent deflationist theory of “true,” therefore, is—and should be—modest. In particular, perhaps only (2) of (1)–(4) from section 17.1 is acceptable to it: a sincere normal assertion of a statement undertakes a commitment to the truth of that statement (and vice versa).

17.3 TRUTH-AScription DEFLATIONISM VS TRUTH-AScription SUBSTANTIVALISM

As just stated at the end of section 17.2, the semantic-descent deflationist theory of “true” is modest. This makes natural the question, “How does this theory bear on the original question raised in section 17.1?” For concreteness, let’s pose it this way: exactly how does the fact that “true” and “false,” as utilized in blind truth and falsity ascriptions, can be replaced *tout court* with substitutional quantifiers or with infinite conjunctions and disjunctions, bear on our original question about what properties, if any, “true” characterizes statements as possessing? The answer is, “In no way at all.” That “true” and “false” can be replaced in their role by generalizing devices that themselves can handle the semantic descent undertaken by “true” and “false” in ordinary language doesn’t mean that a predicate “true” can’t be defined from those generalizing devices, as I indicated in fn. 11. And once such a predicate is available, one can ask (i) whether all and only the statements that fall under that predicate “have something in common”; (ii) whether that something in common can be characterized in some way that’s independent of the use of the word “true”; that’s independent, for that matter, of the generalizing device that “true” is defined in terms of. It should be obvious that an answer to this question isn’t by itself dictated by the deflationist theory of “true.” Notice this point holds of either version of the deflationist theory of “true”: T-schema deflationism or semantic-descent deflationism.

Some philosophers, however, have raised the following concern. There are two ways, they claim, that a concept (and the idioms used to express that concept) can occur in statements. The first is for it to play a purely “expressive” role: it’s employed *only* because it’s needed to express something the expression of which without it is otherwise too awkward or time-consuming. For example, according to T-schema deflationism,

(BLIND₂) Some of what Claire said yesterday is true,

expresses the infinitely long

(INFINITE₂) (Claire said yesterday that s_1 & s_1) or (Claire said yesterday that s_2 & s_2)
... or ...

Other uses of an idiom, however, aren’t expressive, but instead are “substantive”: they *really involve* the attribution of a truth-property to a class of statements.

How does one tell when the use of an idiom in a statement—e.g. “true”—is expressive (as opposed to substantive)? No “bright yellow line” is offered by philosophers who raise this issue. Rather, a number of examples are given where “true” is described as playing an “explanatory role.” It’s this “explanatory role” that indicates that “true” involves the attribution of a genuine “truth-property” to a statement or to several such. Here are

some examples where “true,” it may be suggested, is playing an explanatory role that requires it be read in a property-attributing fashion:

- (CONJUNCTION) For all statements, “A” and “B,” “(A & B)” is true if and only if “A” is true and “B” is true.¹³
- (CONSEQUENCE) All the consequences of true statements are true.
- (BELIEF) We should strive to make our beliefs true.
- (SUCCESS) The success of a true scientific theory can only be explained by its truth.¹⁴

If a philosopher denies that the ascription of “true” to a statement is ever used to attribute a genuine property to that statement, call him a *truth-ascription deflationist*. His opponent is a *truth-ascription substantialist*: she claims that there are truth-ascriptions to statements that do attribute (one or another) truth-property to those statements. One last kind of truth deflationism should be distinguished. This is the previously mentioned deflationist theory of truths. The proponent of the deflationist theory of truths claims that—regardless of whether or not we sometimes utilize truth-ascriptions as the truth-ascription substantialist claims we do—there is (nevertheless) no genuine property that all truths (or even interesting subcollections of truths, such as the truths of empirical science or moral discourse or mathematics, etc.) have in common. These various kinds of deflationism differ importantly in their implications, and should not be confused with one another. In the rest of this section, I explore truth-ascription deflationism and truth-ascription substantialism. Then, in section 17.4, I turn to deflationist and substantialist theories of truths.

Notice, to begin with, the independence of the success of either T-schema deflationism or semantic-descent deflationism from the question of whether these other deflationist theories are true or not. Because CONJUNCTION, CONSEQUENCE, BELIEF, and SUCCESS are blind truth ascriptions, they can be recast (utilizing e.g. substitutional quantifiers) without the truth predicate. CONJUNCTION, CONSEQUENCE, and BELIEF take the follow forms:¹⁵

- (CONJ) $(x)(y)(x \& y \text{ if and only if } x \text{ and } y),$
- (CONS) $(x)(y)(\text{If “}y\text{” is a consequence of “}x\text{,” then, if }x \text{ then }y).$ ¹⁶
- (BEL) We should strive for: $(x)(\text{Believe that }x \text{ then }x).$

¹³ I’m adopting the quotation-mark approach rather than the use of *that*-clauses for this characterization of conjunction for reasons of familiarity and naturalness. Nothing essential to the points to be made is affected. Recall the conventions described in fn. 5.

¹⁴ A related example is this: “It’s because John knew that S was true that John was able to avoid trouble with P.” When it comes to scientific theories, more qualified statements are usually offered, e.g. “The success of a scientific theory can only be explained by its approximate truth.” “Approximate truth” brings up a number of complications that deserve a paper of their own; I can’t undertake discussion of it here.

¹⁵ I take up SUCCESS in section 17.4.

¹⁶ For the sake of naturalness, as with CONJUNCTION, I’m utilizing a quotation-style nominalization rather than a *that*-clause nominalization. In either case, the substitutional quantifier must bind a variable appearing within the nominalization.

Although CONJUNCTION, CONSEQUENCE, and BELIEF being recastable as CONJ, CONS and BEL doesn't in itself show that the truth-ascription deflationist is right, he can still try to use these latter formulations to show that "true" doesn't have (in these statements or in the originals) the explanatory role that the truth-ascription substantialist thinks it has. He can try to show, that is, that the truth of CONJ, CONS, and BEL don't turn on there being a substantial property (of any sort) that true sentences have. In other words, he can try to show truth-ascription deflationism by showing that whether CONJ, CONS, and BEL are true or not doesn't turn on whether the deflationist theory of truths is itself correct or not.

To illustrate how the debate between the truth-ascription deflationist and the truth-ascription substantialist can go, consider

(JOHN) Everything John said is true.

Here, the truth-ascription deflationist can argue, "true" is only being used to indicate an endorsement of what John said: i.e. the focus of JOHN is on endorsing the statements said by John. And so this use of "is true" fits truth-ascription deflationism. But contrast this example with CONJUNCTION, the truth-ascription substantialist may retort. In CONJUNCTION "true" is instead playing a "constitutive role" in the characterization of the logical connective "&." It isn't a matter, that is, of "true" being used to point to a collection of statements that are being endorsed. Instead, CONJUNCTION explains how the ampersand operates on the truth values of the statements that flank it: it yields the value true if and only if each statement is itself true.

The objection to truth-ascription deflationism is, perhaps, even more threatening in the case of BELIEF. Our understanding of belief (many argue) crucially involves its being governed by the *norm* of truth. We strive to make our beliefs *true*. That's why it's incoherent to say sincerely: I believe *p*; but *p* isn't true. It's hard, so the truth-ascription substantialist says, to see how the truth-ascription deflationist view of "true" can capture the role of truth as a norm: how is truth as a norm of belief supposed to come to *nothing more* than its semantic-descent role—its allowing the endorsement of collections of statements? And if it does indicate the special endorsement of *some* collection of statements—"the true ones"—it seems that an extremely important property is being attributed to this particular collection, a property that the truth-ascription deflationist hasn't taken account of because it's a property (of only the true statements) that grounds the truth of BELIEF.

Contrast the role of the true statements in BELIEF with the role of "true" in JOHN as the truth-ascription deflationist describes it. When a speaker assents to

(JOHN) Everything John said is true,

her use of "true" helps provide an endorsement of those statements said by John. But these don't have to be *true* statements. It's simply that the statements John said are endorsed by the speaker of JOHN. The statements indicated by BELIEF, however, have to be taken to be *true* statements. Furthermore, these true statements seem to have an

important property—one that will explain the reason we should strive to make our beliefs true. But how can the truth-ascription deflationist allow such a property to be relevant to the truth of BELIEF?

The truth-ascription substantialist has argued that it's not possible to *interpret* CONJUNCTION and BELIEF without understanding "true" as corresponding to a (substantial) property. But the truth-ascription deflationist can meet this challenge. Consider again,

(CONJUNCTION) For all statements, "A" and "B," "(A & B)" is true if and only if "A" is true and "B" is true,

and its substitutional sibling,

(CONJ) $(x)(y)(x \& y \text{ if and only if } x \text{ and } y).$

Semantic ascent, so obviously present in CONJUNCTION, is absent from CONJ, as indicated both by the vanishing of the "is true" locution, and by the vanishing of the quotation marks. Nevertheless, everything crucial to CONJUNCTION recurs in its substitutional sibling: the interplay between the ampersand and the "and." In CONJUNCTION the ampersand is talked about, and the "and" is used, and in CONJ (true to the nature of substitutional quantification) both are now used. CONJ, as a result, has a decidedly trivial air that's somewhat—but only somewhat—missing from the original. Appearances of triviality lapse, however, when it's realized that both CONJUNCTION and CONJ (recursively) link two different classes of expressions: those with "&" and those otherwise identical but with "and" replacing "&."

What this shows, the truth-ascription deflationist claims, is that "true" is *not* playing a constitutive role in the characterization of "&." What's playing that role (and all that's playing that role) is "and." "True" is needed (in CONJUNCTION but not in CONJ) only because a recursive characterization of "&," when utilizing one kind of quantifier, needs a semantic descent device. But this is inessential to the recursive characterization of "&"—it's an artifact of the choice of quantifier, as CONJ indicates. Therefore, it's not mandatory to invoke truth-properties (truth-"values"), that all statements have, to interpret what CONJUNCTION is telling us: CONJUNCTION is only recursively linking the assertability of &-statements to that of and-statements. This shows, contrary to the claims of many philosophers (e.g. Dummett 1959; Kalderon 1999; Collins 2002) that there is no *incompatibility* between deflationism and truth-conditional theories of meaning.

The truth-ascription deflationist has offered an interpretation of CONJUNCTION differing from the one offered by the truth-ascription substantialist. If the burden of proof is on the substantialist, then the truth-ascription deflationist has won. But what grounds do we have for claiming that the substantialist has the burden of proof? And if she doesn't, why doesn't the debate over CONJUNCTION end in a draw? Why

can't, that is, the truth-ascription substantialist (despite the cogency of a deflationist reading) continue to read CONJUNCTION (and CONJ) as many in fact continue to read them: as describing how the truth-properties of the conjuncts of a statement relate to the truth-properties of the statement itself?

Indeed, the truth-ascription substantialist can so read

(JOHN) Everything John said is true.

It was claimed earlier by the truth-ascription deflationist that when a speaker assents to JOHN her use of "true" provides an endorsement of those statements said by John. And it was claimed that in so endorsing them, they didn't have to be true statements. But this is a misdescription of the situation—so the truth-ascription substantialist can claim—that the discussion of the "endorsement" view of truth at the end of section 17.2 already exposed. JOHN can be *used* to endorse everything John said not because *it* is an endorsement of everything John said but only because if everything John said is true then JOHN is true as well. And certainly such a link between JOHN and the statements it is about can be interpreted in accord with truth being a substantial property. If every statement that John said is indeed true (has the substantial truth-property) then so too is (has) JOHN. So there is a stalemate on JOHN as well.

Now consider BELIEF and BEL, here repeated:

(BELIEF) We should strive to make our beliefs true.

(BEL) We should strive for: $(x)(\text{Believe that } x \text{ then } x)$.

How, the truth-ascription deflationist asks (rhetorically), does BEL go beyond truth-ascription deflationism? It's hard to see how it *could*. After all, what BEL is relying on is the intuition that statements like

I believe that snow is white; but snow isn't white,

are inappropriate. But what makes such statements inappropriate needn't have anything particularly to do with whether or not truth is a property. The inappropriateness of such statements turns, instead, on a misfit between the assertion of a belief with a simultaneous denial of the content of that belief. Indeed, that "truth is a norm of belief" is *itself* a treacherous expression in English that involves a blind truth ascription, namely that if someone believes x , then he should strive such that x be true. But this *is* BEL. The conclusion is this: there *is* a norm governing belief. And, in English, we state that norm using "true," just as we state *all* blind truth ascriptions using "true." But the norm is an internal one about the consistency of assertions of beliefs and assertions of their content; it has nothing to do one way or the other with a presupposition that truth is a property.

The truth-ascription substantialist can agree with this diagnosis of the norm governing belief (and so, she can agree that she has not shown, as she earlier suggested, that the norm governing belief requires a substantial property that truths share and

that grounds that norm). She can still claim, however, that nothing in what the truth-ascription deflationist has said about BELIEF and BEL shows that truth *doesn't* correspond to a property.

Finally, let's turn to

(CONSEQUENCE) All the consequences of true statements are true,

and its substitutional sibling,

(CONS) $(x)(y)(\text{If “}y\text{” is a consequence of “}x\text{,” then, if }x \text{ then }y).$ ¹⁷

Perhaps this case can provide some traction, one way or the other, in the debate between truth-ascription deflationism and truth-ascription substantialism.

Many philosophers describe deduction—when characterized syntactically, as involved with such and such rules—as “explained” by a semantic notion of validity that's based on the concept of truth. Thus, they will describe a rule, *modus ponens* say, as “truth-preserving,” and mean this expression to single out *modus ponens* in a way that does *not* single out the syntactic rules of a non-classical logic. Such philosophers suggest that they are saying something informative about the classical logical principles they accept when they describe them as “truth-preserving.” Such philosophers may even claim that this is where the real difference between truth-ascription substantialism and truth-ascription deflationism arises. The truth-ascription substantialist is interpreting CONJUNCTION and CONSEQUENCE as offering explanations of conjunction and the consequence relation in terms of truth. The truth-ascription deflationist, by contrast, isn't really offering an explanation at all, but only a repetition of what needs explaining (conjunction, consequence) in other language.¹⁸ In response, I want to claim that employing truth-properties in this way in an explanation is only to offer a bogus explanation.

Consider a language with an *arbitrary* set of syntactic rules SY for deductions, apart from one constraint (to be mentioned in this paragraph). If *B* follows from *A* via SY, then the speakers of such a language license themselves as having the right to assert *B* given their right to assert *A* (I'll write this as: $A \vdash B$). Now also imagine that they have a notion of “true” for which $\text{True}A \vdash A$, and $A \vdash \text{True}A$ hold. It follows immediately that if $A \text{ SY } B$, and $\text{True}A$, then $\text{True}B$.

The point is this: given *any* consequence relation (defined by an arbitrary set of syntactic rules), if, for any statement *S*, $\text{True}S$ is a consequence of *S* (and vice versa), then that consequence relation is “truth-preserving.” Conclusion: reading CONSEQUENCE

¹⁷ Again, for the sake of naturalness, I'm utilizing a quotation-style nominalization rather than a *that*-clause nominalization. In either case, the substitutional quantifier must bind a variable appearing within the nominalization.

¹⁸ My thanks to Michael Glanzberg for posing the substantialist objection in this particular form.

as the truth-ascription substantialist wishes no more singles out *modus ponens* (and explains it) then it does any *arbitrary* inference rule.

This refutation of the claim that substantialist readings of CONSEQUENCE and CONJUNCTION in terms of truth-properties provide better explanations than deflationist readings of them, however, leaves unaffected the possibility of nevertheless giving them a substantialist reading. That is, the truth-ascription substantialist can agree that considerations of truth-transmission (validity) don't force the choice of any particular logic. Granting that, however, and granting a particular consequence relation SY, she can claim that such a consequence relation between statements describes how the possession of a (substantial) truth-property by certain statements relates to the possession of that property by other statements. It may indeed be the case that certain truth-properties that can be specified may prove to be incompatible with certain consequence relations. But this hardly shows that CONSEQUENCE *must be* read as the truth-ascription deflationist wants to read it. We again have a stalemate.

I said earlier that there being equally acceptable readings of any particular locution containing the word "true" yields a draw in the debate between truth-ascription substantialism and truth-ascription deflationism only if there is no burden of proof on the truth-ascription substantialist. One may suggest, however, that there is such a burden of proof, although its force doesn't emerge until it becomes likely that every locution can be read compatibly either with truth-ascription deflationism or with truth-ascription substantialism. Now that a truth-ascription deflationist reading has been given for CONJUNCTION, JOHN, CONSEQUENCE, and BELIEF, the following challenge can be issued to truth-ascription substantialism. If *every* occurrence of "true" in any locution is compatible with either a deflationist reading or a substantialist reading, why should anyone believe that there is a substantial property that's being attributed to truths? *In general*, that the use of a kind term holding of something corresponds to an ascription of a substantial property to that something surely requires the purported substantial property to play a role in the evaluation of the truth or falsity of at least *one* statement in which that kind term appears. If this demand (on substantial property attributions) is right, the truth-ascription substantialist has the burden of proof: at least one statement S in which "true" appears must have a truth value that turns on whether the statements "true" is attributed to by S have the substantial property in question. Luckily for the truth-ascription substantialist, there is such a statement (more than one, actually), as I now show. In doing so, I show in what follows how issues about truth ascriptivisms connect to issues about deflationary and substantialist theories of *truths*.

17.4 SUCCESS AND TRUTH

Consider SUCCESS, here repeated:

(SUCCESS) The success of a true scientific theory can only be explained by its truth.

SUCCESS involves complications both because scientific theories aren't mere collections of statements, and because the idiom of explanation requires nominatives. We say, for example,

The movement of the tides is explained by the orbit of the moon,

and not

The tides move (in such and such ways) is explained by the moon orbits the Earth (in so and so ways).

For illustrative purposes, however, I idealize scientific theories as collections of statements, and I similarly introduce a conception of "is explained by"—parasitic on the ordinary notion of explanation—that grammatically operates as a sentence connective. Having done so, the semantic-ascent deflationist (or the T-schema deflationist) can handle SUCCESS as CONJUNCTION was handled:

(SUC) (x)(If "x" is a true scientific theory, then: "x" is successful is explained by x).

Both SUCCESS and SUC have a trivial feel to them that seems to rule out the possibility of their being true. SUCCESS tells us that it's the fact that true scientific theories are true that explains why they're successful. But it's hard to see why a truism—*true* scientific theories are *true*—should explain *anything*. (Putnam 1978 first raised the issue of the role of truth in the explanation of scientific truths. See Davidson 1990; Field 1986; 1994, and Kitcher 2002, among others, for further discussion.)

What seems to provide the needed explanation is something that neither SUCCESS nor SUC talk about: the correspondence relations between the terms of true scientific theories and the objects those theories are about. Consider for instance a theory MO of molecular biology, one that enables biologists and their technical assistants to manufacture new organisms. Surely the success of such biologists and their assistants is due to MO being true, where its being true is its being true *of* items that the biologists and their assistants manipulate with the aid of MO. What seems to be missing from SUC and SUCCESS is a description of the needed correspondence of MO with the objects in question that biologists have been so successful with.

SUCCESS and SUC only seem to focus on this: There is a set of statements that are true. Call these "the truths" (TR). SUCCESS and SUC claim that it's because the statements of a successful true scientific theory are contained in TR that such a theory is successful.

Here's another example of how the focus of SUCCESS and SUC seems to fall short. Imagine that FUR is a collection of ordinary statements about the arrangement of furniture in a room. TR, also imagine, contains a description of the *actual* arrangement of furniture in a room. If FUR describes the location of the furniture in that room correctly,

if the sentences of FUR are *in* TR, then in using FUR, we'll maneuver successfully about the room; if FUR isn't true, we won't maneuver successfully about the room.¹⁹

With both MO and FUR, it's not just a set of truths being contained in another set of truths that seems to be providing the needed explanation of success; it's that the truths in question *describe* particular correspondence relations that provide the needed explanation of success. But this latter fact seems to go beyond the content of SUC and SUCCESS.

One can try to counter the triviality charge with this idea: whether SUCCESS and SUC are true (or false) of a set of true statements turns on *how* those statements are true, and how that relates to whatever their success comes to. If a statement's being true *amounts to* how it describes the layout of furniture, and if the success of that theory turns directly on how well it guides us as we maneuver around the furniture, then this is stated by SUC and SUCCESS because the success of any true scientific theory turns directly on what the theory *says*, and on what it says being so. Thus, what MO and FUR say *does explain* their success.

Nevertheless, the feel of triviality (*true* scientific statements are contained in the set of *truths*) is hard to shake off. And because of this apparent triviality, the explanation of success provided by SUCCESS (and SUC), therefore, seems to be available for *any* set of true sentences whatsoever. After all, *any* set of true statements is contained in the truths. This is the key to why the explanation in question isn't trivial: a set of true statements being contained among the truths *doesn't* explain the success of *every* set of true statements. There are cases where, even though a set of truths is contained in the set of truths (as they must be), the success of that set of truths must be explained in some other way.

Here is an illustration. Consider the following nominalist picture of mathematical truth. There are no mathematical objects, but mathematical statements are true (and false) nevertheless. On this view, what *makes* true mathematical statements true is their (permanent) indispensability, say, to scientific practice. If this nominalist+indispensabilist description is the correct picture of mathematical truth, then the versions of both SUC and SUCCESS (when "scientific" is replaced with "mathematical") are *false*. It isn't the *content* of true mathematical theories (what they apparently *say* about mathematical entities) that explains the success of those theories; what explains the success of true mathematical theories are the details of how and why they are permanently indispensable to scientific practice. The (permanent) indispensability of any such true mathematical theory is also *why* such a mathematical theory is true. But

¹⁹ One important complication is being set aside. SUCCESS is misleading because it gives the impression that all that's relevant to the success of a true theory is the theory itself. But that's never true. MO enables biologists to manufacture certain biological entities successfully not just because *it* is true, but because of other truths about biologists, their tools, and how these can be brought to bear on the entities that MO is about. So too, those who maneuver about a room filled with furniture do so successfully not only because of the truth of FUR but (among other truths) those truths about their own bodies and truths about the relations of their bodies to furniture in the room. Taking account of these subtleties won't change the course of the argument.

its truth therefore *doesn't* explain its success. For mathematical theories, where correspondence is absent, instead of SUCCESS, we have

(SUCCESS*) The success of a true mathematical theory can only be explained by its (permanent) indispensability.

What is the upshot of the somewhat complicated dialectic that's been pursued in this section and the last one? First, SUCCESS and SUC are statements *the truth of which requires a correspondence between the truths they are about and objects in the world*. The way SUC and SUCCESS manage this, when neither speaks explicitly of a correspondence, is that it's only when a statement *does* involve a correspondence that its *mere* truth can explain its success. This is the lesson that SUCCESS* teaches: if mathematical truths are unaccompanied by correspondence, *something else* is needed to explain what makes mathematical truths successful.²⁰ This means the truth-ascription substantialist is right and the truth-ascription deflationist is wrong.

Recall from section 17.3 the distinction between “expressive” and “substantive” uses of “true.” Expressive uses of “true” are those where “true” is purportedly being employed only because relinquishing its use results in otherwise awkward or time-consuming expressions. Substantive uses of “true” really involve the attribution of a truth-property to a class of statements. The distinction between the uses of “true” that can be eliminated and those that can't is a distinction that many philosophers have supposed to match the distinction between uses of “true” compatible with truth-ascription deflationism and uses that aren't—but we can see now that these distinctions don't match at all.

(SUCCESS) The success of a true scientific theory can only be explained by its truth, involves an expressive use of “true”: SUCCESS can be replaced by

(SUC) (x)(If “x” is a true scientific theory, then: the success of “x” is explained by x).

Nevertheless, both of these claims do attribute a correspondence property to true scientific theories. Consequently, there is a statement in the neighborhood of SUCCESS that in turn poses a direct threat to the deflationary theory of truths, the denial that there is a property P that all truths have. Consider:

(GLOBAL SUCCESS) The success of *any* set of true statements can only be explained by the truth of those statements.

²⁰ This makes correspondence truth different from the other truth-properties: coherence, utility, etc. The latter truth-properties require explanations of success in terms other than truth itself. Only when a true statement describes items that its terms correspond to can mere truth play an explanatory role in that statement's success (apart from, of course, the points made in fn. 19—ones that every explanation of the success of a set of statements must take account of).

If GLOBAL SUCCESS is true, then the deflationary theory of truths is false. If nominalism, as earlier described, is true then GLOBAL SUCCESS is false: to explain the success of true mathematical theories, one must employ something like SUCCESS*, which is *not* a specification of GLOBAL SUCCESS to collections of true mathematical statements.

In any case, what these considerations about SUCCESS, SUCCESS*, and GLOBAL SUCCESS show is that there are statements we can formulate whose truth or falsity turns on the *particular* truth-property they attribute to a set of true statements. If a set of true statements involves (say) referring terms the relata of which are described by those true statements, and if various sorts of success induced by the use of those true statements are directly linked to our utilization of the descriptions of the relata given by those statements, then a *specification* of GLOBAL SUCCESS to the collection of those statements is true. If, however, the property that a set of true statements has is one that's like the suggested property of true mathematical statements given earlier—e.g. that such are permanently indispensable to such and such epistemic practices of ours—then a specification of GLOBAL SUCCESS to them is false; instead, something like SUCCESS* is true of them.

In what ways, of course, indispensability in turn connects to truth is open to further analysis. I've offered as a suggestion that, in the case of (pure) mathematical statements (if they are as the nominalist describes them), permanent indispensability yields truth:

(INDIS TRUTH) A permanently indispensable mathematical statement is true.

Notice that INDIS TRUTH may be expressed by substitutional quantifiers:

(INDIS) (x)(If "x" is mathematical, and permanently indispensable, then x).²¹

17.5 TRUTH-PROPERTY PLURALISM

Many philosophers in the last decades of the twentieth century thought that a description of the role of "true" in ordinary language impels the conclusion that truth ascriptions aren't attributions of a property to the statements so ascribed. But the distinctions of section 17.3 between deflationist theories of "true," truth-ascription deflationism, and the deflationist theory of truths show that this discussion-area must

²¹ Some philosophers have argued for:

(ASYMMETRY) It is true that S because S, but not vice versa.

My analysis of ASYMMETRY parallels my analysis of SUCCESS. ASYMMETRY is true of a statement S when the terms of S (all) involve correspondence relations to objects, but not otherwise.

be reconfigured. Although semantic-descent deflationism is correct, its truth doesn't bear on either truth-ascription deflationism or on the deflationist theory of truths. In point of fact most, but not all, truth ascriptions can be construed either in accord with truth-ascription deflationism, or against it. Some statements, however, such as SUCCESS, attribute a correspondence-property to (sets of) truths, contrary to truth-ascription deflationism.

Finally, whether in fact all true statements do share some property such as a correspondence property (whether or not, that is, a deflationist theory of truths is correct) turns on the kinds of consideration that an earlier tradition in philosophy presumed were involved: the nature of the relata (if any) of the terms of the statements described as true and false.

Mathematical nominalism is a live issue to be decided in part on metaphysical grounds.²² If it's right, then the truth of mathematical statements does indeed turn out somewhat "epistemic" in character: it's the permanently indispensable role of mathematical statements in our science that forces their truth. This isn't solely a matter of "how the world is," but is equally a matter of the nature of the statement-vehicles we must use to describe that world. On the other hand, such a claim needn't extend to all statements we (provisionally) take to be true. Some statements may indeed be true only because of a correspondence between the way they describe things as being, and the way those things themselves are. The conclusion to be drawn is that the deflationist theory of truths is right, not because "true" is never used to ascribe a property to statements, but because different statements are true for different reasons. There is no property—relational or otherwise—that can be described as what all true statements have in common (other than, of course, that they are all "true").

Some philosophers have drawn a different conclusion (e.g. Wright 1992; Lynch 2004): we engage in different kinds of discourse, for example about mathematics, about empirical objects, about fictions, and about morality. These discourses are segregated from one another; and these philosophers claim that the truth predicate, when applied to statements in any of these discourses, corresponds to a truth-property shared by the statements in that discourse, but not necessarily shared by statements from other discourses. For example, although the truth predicate, in an empirical-object discourse, can involve a correspondence-truth-property, when that truth predicate is instead attributed to mathematical discourse, it will not.

This pluralist view of truth-properties, however, faces a severe difficulty because of the simple fact that we do *not* engage in different kinds of discourse. Moral discourse, for example, involves statements that mix together moral and empirical vocabulary; indeed, it's scarcely possible to imagine the point of statements containing only moral vocabulary. In the same way, one finds mathematical vocabulary occurring in statements along with empirical vocabulary: our most indispensable and valuable empirical statements

²² This is in part because the soundness of the Quine-Putnam Indispensability argument is also a live issue that the status of mathematical nominalism turns on.

routinely contain both mathematical and empirical vocabulary. Even ordinary statements about fiction inextricably involve nonfictional vocabulary:

Mickey Mouse was invented by Walt Disney.

The truth-property pluralist correctly notices that the terms of true statements vary: some refer and some don't. If a true statement contains only referring terms, it seems a correspondence-truth view applies; if a true statement doesn't contain such terms, this is not so. The truth pluralist is wrong, however, in thinking that statements can be segregated according to "subject-matter" in different discourses. At best what's true is that terms can be so segregated; but this is cold comfort for philosophers committed to there being various truth-*properties* because truth-properties must be applicable to whole sentences. This is too crude, because the vast majority of valuable sentences are each composed of terms from several different vocabulary areas.

17.6 TRUTH BEING SUBSTANTIAL VS "TRUE" NOT BEING CONSERVATIVE

There is one loose end the discussion of which illustrates some of the lessons of this chapter. Some philosophers have brought certain formal results to bear on the debate between truth-property deflationists and truth-property substantialists (e.g. Shapiro 1998; Ketland 1999). If a standard formal language can express a small amount of arithmetic, then no sufficiently-rich consistent theory in that language can express a truth predicate for it. Instead, if, in TH_2 (in another language) we characterize the syntax of TH_1 (in the first language), and define a truth predicate for TH_1 , we can express more and prove more in TH_2 than we can in TH_1 . This result is Tarski's.

A related result is this: consider an axiomatization A of a small bit of arithmetic that includes the principle of induction. Now consider a new axiomatization A^* that includes A plus a truth predicate for A that can occur within the induction schema. A^* is strictly stronger than A in the following sense. There are sentences in the language of A which can be proved from A^* but not from A .

How do these results bear on the debate between truth-property deflationists and truth-property substantialists? What, in other words, is the connection between "true" failing to be conservative in formal contexts, and "true" therefore having to correspond to a "substantial" property? We have to speculate because those raising the relevance of these formal considerations have never explicitly argued for the connection. Here is a suggested "argument." If "true" is *just* a device of endorsement, then its conceptual content can't come to more than the statements it's used (on an occasion) to endorse. How, then, could adding such a predicate to an axiomatization about numbers add content so that new results about those numbers could be proved that couldn't be proved before? These considerations impel the conclusion that "true" can't merely be a device of

endorsement, and therefore that “true” must correspond to a substantial property, some concept of which supplies the content that explains why adding it to an axiomatization of numbers isn’t conservative.

We now see several mistakes in this line of thought. I’ll indicate some. There is first the mistaken assimilation of how a truth predicate is used (even when such utility is the reason for the presence of the device in ordinary language) to what its properties actually are. A truth predicate isn’t a device of endorsement, direct or indirect, although it’s commonly used that way. What it is, is a *device of semantic descent*. A device of semantic descent *needn’t* be proof-theoretically conservative, nor should that be expected. Gödelian and Tarskian results turn directly on what can and can’t be expressed in a formalism, and those results also turn on how expressive resources are and aren’t segregated in different languages. A device of semantic descent (roughly speaking) moves those expressive resources from one language to another, and this is why such a device—in the right circumstances—won’t be conservative.

I’ve speculated on a particular line of thought connecting the non-conservativeness of the truth predicate in certain formal settings to a challenge to truth-property deflationism. But *any* such approach to a challenge to truth-property deflationism must turn on assuming that it isn’t possible for a truth predicate to have genuine content that it can contribute to an axiomatization (i.e. that it’s a device of semantic descent) and yet for such to fail to implicate a property. There is a failure, that is, to distinguish the very different claims of a deflationist theory of “true” from that of truth-property deflationism and from a deflationist theory of truths. Keeping clear the distinctions between these very different kinds of deflationist theory “about” truth is the most important lesson of this chapter. For in this way one can understand why a device of semantic descent that’s a predicate is *open* to either corresponding to a substantial shared property or not so corresponding.²³ Being a device of semantic descent, and being non-conservative in formal contexts as a result, raises no challenge to deflationist positions about truth.

17.7 SUMMARY OF RESULTS

In this chapter, I’ve made a number of important distinctions that have been—to various degrees—confounded in earlier literature. I’ve distinguished two versions of deflationist theories of “true,” and distinguished these in turn from a deflationist theory of truth-*ascription*, as well as from a deflationist theory of *truths*. I’ve further suggested that the semantic-descent deflationist theory of “true” and the deflationist theory of truths are both correct, although the considerations that must be brought to support or attack these different deflationist theories are largely independent of one another. I’ve

²³ In particular, understanding “true” *purely* as a device of semantic descent is quite compatible with an interpretation of (certain) Gödel sentences as *true* on the grounds of a correspondence interpretation of what they say.

denied, however, the view that our attributions of “true” to collections of statements are invariably ones that *don’t* attribute (say) a correspondence property to those truths. (Some statements do attribute such a property, I’ve claimed, and I’ve claimed further that statements that are suitably restricted with respect to the statements they attribute this property to are *correct*.)

A careful understanding of deflationist theories of “true” requires recognizing certain roles of “true” and “false” in ordinary language; this in turn requires distinguishing how these words are useful from what linguistic properties they have. “True,” in particular, is not a device of endorsement or a device of generalization. A similarly careful understanding of the deflationist theory of truths requires recognizing that any such theory is about truths and not about the properties of, nor the use of, the word “true” in ordinary language or in formal languages. Whether truths share some relational property or other—e.g. correspondence, coherence, etc.—or whether they don’t, turns on issues about the nature of the grounds of truths: what it is, if anything, that *makes* statements true.

The word or concept “true” can have substantial conceptual content without that content being an attribution of a substantial property in the sense that the old philosophers (who debated over whether truth involved correspondence to a “mind-independent reality,” or instead involved inextricable “epistemic content”) had in mind. To recognize “true” is a device of semantic descent is certainly to attribute conceptual content to *it*. But that content does *not* implicate the further content that what is true must share some substantial property or other.²⁴

REFERENCES

- Azzouni, J. (2001). “Truth Via Anaphorically Unrestricted Quantifiers,” *Journal of Philosophical Logic* 30: 329–54.
- Azzouni, J. (2006). *Tracking Reason: Proof, Consequence, and Truth*. Oxford: Oxford University Press.
- Blackburn, S. and K. Simmons (eds.) (1999). *Truth*. Oxford: Oxford University Press.
- Collins, J. (2002). “Truth or Meaning? A Question of Priority,” *Philosophy and Phenomenological Research* 65: 497–536.
- Davidson, D. (1990). “The Structure and Content of Truth,” *Journal of Philosophy* 87: 279–328.
- Dummett, M. (1978). “Truth,” in M. Dummett, *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 1–24.
- Field, H. (1986). “The Deflationary Conception of Truth,” in G. Macdonald and C. Wright (eds.), *Fact, Science and Morality*. Oxford: Oxford University Press, 55–117.
- Field, H. (1994). “Deflationist Views of Meaning and Content,” *Mind* 103: 249–84.
- Field, H. (2001). *Truth and the Absence of Fact*. Oxford: Oxford University Press.
- Gupta, A. (1993). “Minimalism,” *Philosophical Perspectives* 7: 359–69.

²⁴ My thanks to Bradley Armour-Garb, Michael Glanzberg, and Douglas Patterson for help on earlier versions of this. My thanks to Dan Greco for pointing out that one of my rasher claims needed moderation (oral communication on November 30, 2012).

- Horwich, P. (1998). *Truth*, 2nd edn. Oxford: Oxford University Press.
- Kalderon, E. (1999). "The Transparency of Truth," *Mind* 106: 475–97.
- Kitcher, P. (2002). "On the Explanatory Role of Correspondence Truth," *Philosophy and Phenomenological Research* 64: 346–64.
- Leeds, S. (1978). "Theories of Reference and Truth," *Erkenntnis* 13: 111–29.
- Lynch, M. P. (2004). "Truth and Multiple Realizability," *Australasian Journal of Philosophy* 82: 384–408.
- Ketland, J. (1999). "Deflationism and Tarski's Paradise," *Mind* 108: 69–94.
- Künne, W. (2003). *Conceptions of Truth*. Oxford: Oxford University Press.
- Putnam, H. (1978). *Meaning and the Moral Sciences*. London: Routledge & Kegan Paul.
- Quine, W. V. (1970). *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice Hall.
- Ramsey, F. P. (1927). "Facts and Propositions," *Proceedings of the Aristotelian Society Supplementary Volumes* 7: 153–70.
- Russell, B. (1912). *Problems of Philosophy*. Oxford: Oxford University Press.
- Schantz, R. (ed.) (2002). *What is Truth?* Berlin: Walter de Gruyter.
- Shapiro, S. (1998). "Proof and Truth: Through Thick and Thin," *Journal of Philosophy* 95: 493–521.
- Tarski, A. (1983). "The Concept of Truth in Formalized Languages," in J. Corcoran (ed.), *Logic, Semantics, Metamathematics*. Indianapolis: Hackett, 152–278.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

CHAPTER 18

TRUTH IN FICTIONALISM

ALEXIS BURGESS

WHAT does realism about an arbitrary subject-matter have to do with truth? Michael Dummett says, “Everything”; Michael Devitt says, “Hardly anything.”¹ Both answers are reflected in ongoing debates between self-styled realists and anti-realists in metaphysics, metaethics, the philosophy of mathematics, and beyond. On the one hand, realism about Fs is standardly understood as the view that Fs objectively exist.² On the other hand, error theory, nonfactualism, fictionalism, and other forms of opposition to realism are normally articulated and differentiated using the notions of truth and falsity. Given its preoccupation with the limits of literal representation, fictionalism can seem especially ensnared in semantics and/or the theory of mental content. Be that as it may, the present chapter aims to establish that there remains an important sense in which the fictionalist gambit doesn’t essentially have anything to do with truth or falsity.

I say “an important sense” just because there are different ways to operationalize the idea that fictionalism isn’t an alethic position, as we’ll put it. One way would be to say that some versions of fictionalism are compatible with nihilism about truth: the view that (the property of) truth simply doesn’t exist. But if Quine is right about the ontological commitments of predications, then this thesis will be trifling. Even the proposition that “Snow is white” is true won’t count as alethic in the stipulated sense (just as there can be red houses in the absence of redness). So let’s say instead that a thesis is alethic when it entails that there are truths. Thus the central claim of the chapter, defended in section 18.2, is that many recognizably fictionalist positions are compatible with nominalism about truths: the view that nothing whatsoever is true.³

¹ See e.g. Dummett (1982); Devitt (1984).

² Where “F” stands in for an arbitrary count noun. The difficulties of adapting this scheme to define scientific realism, for example, are of course part of the motivation for appealing to truth.

³ We’ve seen that nihilism about truth does not entail nominalism about truths. Conversely, nominalism about truths does not entail nihilism about truth: one might simply think that nothing instantiates the property of *being true*. The two ‘isms are logically independent. Devitt would describe both nominalism and nihilism as forms of anti-realism about truth(s), but what we call them doesn’t matter much to me.

It will also be important to distinguish compatibility with nominalism or nihilism from the meta-linguistic feature of being impossible to formulate without using the notion of truth. Deflationists⁴ typically think that nothing is alethic in this third sense, yet they do usually think that many uses of “true” commit us to the existence of truths. One reason to formulate our thesis in terms of nominalism is therefore just to keep deflationists interested. The instance of fictionalism adumbrated in the third and final section of the chapter—fictionalism about truth(s)—can actually be construed as a kind of deflationism to rival minimalism or prosententialism. As we’ll see, the coherence of this position is more directly threatened by the possibility that fictionalism in general entails the existence of truths than by the possibility that fictionalism can’t be formulated without using the notion of truth.

Let’s begin by surveying the many faces of contemporary fictionalism. Organized around a couple of basic, binary distinctions, the catalogue in section 18.1 should prove useful even to those with limited interest in the rest of the chapter.

18.1 TAXONOMY

A *version* of fictionalism, as I’ll use the italicized term, is given by an explicit definition of the phrase “fictionalism about Fs.” Since there are almost as many versions of fictionalism as there are fictionalists, we’ll do well to mark alternative definitions with distinctive adjectives (e.g. hermeneutic prefix fictionalism). An *instance* of fictionalism can then be derived by substituting some count noun for “F” in any given version of fictionalism (e.g. fictionalism about universals). Not until section 18.3 will we really be concerned with the details of any particular instance of fictionalism, though different instances will come up as examples in the present section, to help clarify the terms in our taxonomy of current versions of fictionalism.

Early expositions of fictionalism about Fs understood the view to begin with a commitment to the thesis that sentences of the form “b is an F” are systematically false. In other words, fictionalism traditionally began in a kind of error theory. This is presumably part of the reason fictionalism is so called: on a natural view, fictions are false. But fictionalism gradually divorced itself from error theory, largely for the following two independent reasons. In the first place, the “natural view” of fiction doesn’t hold up under scrutiny. Certainly many of the sentences found in many works of fiction are false; but some such sentences are true. Just think of simple, existentially-quantified sentences from realistic fictions, or sentences from historical fictions describing the historical facts. (One can even imagine a story consisting entirely of truths, under the supposition that the author got reality right by accident.) In the second place, Hartry Field’s pioneering fictionalism about arithmetical entities was often assimilated to the contemporaneous constructive

⁴ See Azzouni (ch. 17 in this volume).

empiricism of Bas van Fraassen.⁵ Since the latter was officially neutral on the truth or falsity of sentences (ostensibly) about unobservables, and since it's easier to agree with someone who says less, many future fictionalists came to refrain from claims of falsity.

More importantly for our purposes in section 18.2: the main motivation for error theory in any given domain is typically nominalism about the putative entities of that domain. Versions of fictionalism founded on error theory can therefore be straightforwardly reconceived as resting directly on nominalism (or agnosticism motivated by the possibility of nominalism)—which would seem to be a purely metaphysical thesis. Instead of saying that every sentence of the form “b is an F” is false, the fictionalist can simply say that there aren't any Fs to speak of. But in order to verify that this reformulation comes at little cost, we'll have to say more about the content of contemporary fictionalism, over and above nominalism or error theory.

One distinction worth making at the outset is that between hermeneutic and revolutionary versions of fictionalism, to use the language popularized by Burgess and Rosen. As the terms suggest, the distinction they mark is just that between description and prescription. The hermeneutic fictionalist tells us how certain things happen to be, whereas the revolutionary fictionalist recommends that we make things a certain way—usually the way his hermeneutic counterpart takes them to be. (If things are already as the revolutionary fictionalist recommends we make them, so much the better for him; his advice is vanishingly easy to take.) We can illustrate the hermeneutic/revolutionary distinction by introducing what I'll call *prefix fictionalism*.

As a first approximation: the hermeneutic prefix fictionalist about Fs thinks that “what we really mean” when we say anything⁶ of the form (A) “. . . F . . .” is the corresponding thing of the form (B) “According to the F-fiction, . . . F . . .” His revolutionary counterpart advises that, whatever we currently mean by the former, we decide to mean the latter from now on. (At the risk of duplicity, we might leave our fictionalization prefixes tacit to facilitate conversation with the uninitiated.) The F-fiction is essentially just the realist's theory of the relevant subject-matter, and thus will at least entail that there are Fs. So, even if there aren't any Fs, many of the things we (currently or come to) mean when we use “F” may well be true, according to the prefix fictionalist. Compare (e.g.) “Linda is a witch” and “According to the witch-fiction, Linda is a witch.” This is the primary motivation for any variation on prefix fictionalism (as opposed to the “pretense” versions we'll discuss shortly): to have us speak truly,⁷ despite the threat of error.

What do prefix fictionalists mean by the phrase “what we really mean”? Presumably they shouldn't be understood as saying that A and B are simply synonymous, as that

⁵ Both forms of anti-realism emphasize virtues, like utility, that a theory can exhibit without being true, while steering clear of instrumentalism and other forms of non-cognitivism about their respective domains. Moreover, van Fraassen's notion of acceptance is often analogized to the notion of pretense or make-believe at play in some versions of fictionalism.

⁶ Or, more precisely, to avoid the obvious regress: anything unprefixes.

⁷ This is of course exactly the sort of grammatically expedient use of the word “truly” that deflationists emphasize.

would render the fictionalization prefix redundant, undermining the possibility of speaking truly in the face of error. Nor should we take them to mean that the semantic content of A is somehow (asymmetrically) given by B, since compositionality would suggest the opposite. A more charitable reading would distinguish speaker-meaning from semantic meaning, attributing to the fictionalist the idea that what we speaker-mean by A is just what B means as a matter of semantics. In other words, we're just being elliptical when we use A instead of B. This suggestion certainly respects the spirit of revolutionary prefix fictionalism, though wrinkles admittedly remain in its application to hermeneutic versions of the view.

If the semantic meaning of B is given (in part) by the semantic meaning of A, then presumably, if we are capable of speaker-meaning the former we should be capable of speaker-meaning the latter. But if we are capable of speaker-meaning the semantic meaning of A, why wouldn't we use A to do so? One response to this challenge would allow that young children often do use the simple sentences to express their semantic meanings, but insist that adults are usually less naïve. The problem with this response is that ordinary adults typically won't acknowledge that they're tacitly using fictionalization prefixes; and who are we to tell them what they really speaker-mean? (The analogous notion of speaker reference, note, is usually motivated by cases in which speakers have more or less explicit referential intentions.) Objections like these can be used to motivate other forms of fictionalism, but I don't think they show that we've mischaracterized hermeneutic prefix fictionalism. As we've seen, an appeal to something like speaker-meaning or ellipsis is needed just to get the view off the ground.

Some prominent variations on prefix fictionalism are even less straightforward. Gideon Rosen's so-called modal fictionalist, for example, appends his prefix to *paraphrases* of the target sentences rather than the sentences themselves. In this instance, the target is metaphysical modal discourse, and the relevant paraphrase of a given sentence is its standard analysis in terms of possible worlds (construed à la Lewis, in order to avoid canonical objections against the ersatz). Appealing to paraphrase affords the prefix fictionalist more freedom in dealing with the semantic issues discussed above. Rosen proposes that modal claims are materially equivalent to their prefixed possible-worlds analyses; one could even coherently maintain that the equivalence is one of synonymy.⁸ The so-called modal fictionalist can therefore say (it's a fact) that there could have been blue swans, which can easily lead to terminological confusion.

How should we understand the relationship between simple prefix fictionalism and this variation involving paraphrase? One could think of the former as a special case of the latter, where the paraphrase function simply maps each sentence to itself. Alternatively, one could think of the latter as offering a semantic addendum to the former. With respect to Rosen's topic, the second option would amount to a reading on which we combine simple prefix fictionalism about *possible worlds* with an account of the truth-conditions (or speaker-meanings, or whatever) of statements involving the

⁸ Though there is some reason to worry that these claims are to blame for the problem described in Brock (1993).

notions of metaphysical possibility and necessity. Looking forward to the next section, we should note that the two components of this reading of paraphrase fictionalism can be separated, and that the semantic addendum about modal discourse has very little to do with the nominalist motivations for fictionalism about possible worlds. So, even if it were to turn out that paraphrase fictionalism is ineluctably alethic, there might remain an interesting sense in which “pure” prefix fictionalism is compatible with nominalism about truths.

However this may be, any survey of the varieties of fictionalism would be incomplete without some treatment of what I’ll call *pretense fictionalism*. As against the prefix fictionalist, the pretense fictionalist grants that we really mean what we *seem* to mean when we use sentences of the relevant discourse. But the pretense fictionalist thinks that we should (in the revolutionary case) or do already (in the hermeneutic case) take an attitude other than belief toward the contents of the sentences from the discourse that we are inclined to assert. Pretense fictionalists deserve to be called “fictionalists” insofar as they think the attitude we do or ought to take is much like the one we routinely take toward propositions constitutive of fictional stories. If nominalism about Fs or error theory about F-discourse is right, then the pretense fictionalist about Fs will think and say many things that are literally false (in contrast to the prefix fictionalist), but he’ll merely pretend or “make-believe” what he thinks and says.

Prefix and pretense fictionalism would amount to the same thing if it turned out that pretending such-and-such just *is* believing such-and-such is true according to the pertinent fiction. But this is an unpopular view, and even those attracted to it can grant that prefix and pretense fictionalism are at least notionally distinct. How they differ from realism about the relevant topic (and thus from each other) can be summarized in the following pair of mantras. Pretense fictionalism: same content, but a different attitude toward that content. Prefix fictionalism: different content, but the same attitude toward that content.

Having expressed the two views at this level of abstraction, we can now see how a third, equally influential form of fictionalism—what Steve Yablo has called “figuralism”—might fit in to the framework we’ve been developing. Yablonian fictionalism effectively incorporates both combinations of content and attitude described above. A given sentence from the target discourse is associated with two different contents: its literal semantic content, which the fictionalist does not believe but perhaps make-believes, and its figurative content, which the fictionalist actually believes. To borrow an illustration from Yablo: consider metaphorical uses of the phrase, “He’s on fire.” When a spectator at the NBA finals says this of Big Baby Davis, he obviously doesn’t believe the literal content of his utterance. (Otherwise he’d presumably look for an extinguisher or call for help.) At most, the spectator make-believes this content en route to believing its figurative content: something to the effect that Big Baby Davis has just made three shots in a row. How exactly the figurative content of a given claim gets fixed has, understandably, been the topic of much discussion.⁹

⁹ Jason Stanley (2002) complains that compositionality will be violated.

By my accounting, then, there are eight main versions of fictionalism—hermeneutic and revolutionary variations on each of the four views in bold below, organized with respect to the distinctions in content and attitude we’ve been discussing:

<i>Attitude:</i>	Make-Belief	Belief		
<i>Content:</i>	Literal	Figurative	Literal + Prefix	Paraphrase + Prefix
<i>Version:</i>	Pretense	Yablo-Style	Prefix	Rosen-Style

Note that Yablo actually straddles the border with pretense fictionalism, and that realism could be represented by a diagonal connecting the attitude of Belief to Literal content. No doubt there are other, interestingly different versions of fictionalism to be found in the literature (especially pre-1980),¹⁰ but I will restrict our attention to these eight for present purposes. Even if a version we’ve omitted turns out to be essentially alethic, our conclusions in the following sections will still hold. Next I’ll argue that revolutionary and hermeneutic pretense and prefix fictionalisms are compatible with the view that nothing is true. Then, in section 18.3, I’ll advance a revolutionary prefix fictionalism about truth(s).

18.2 WHAT’S TRUTH GOT TO DO WITH IT?

Strictly speaking, *revolutionary* fictionalism clearly isn’t an essentially alethic thesis, since it isn’t a thesis at all. It’s rather a recommendation or piece of advice, perhaps best expressed by a hypothetical imperative, “If you want to continue using F-discourse but avoid ontological commitment to Fs, then construe what you say thusly . . .” Imperatives aren’t alethic in our operationalized sense, because they aren’t consistent or inconsistent with *anything*. Of course, two imperatives can be compatible (or incompatible) in the practical sense of being mutually satisfiable (unsatisfiable). But it would be a category error to say that an imperative is compatible with an indicative—in the relevant, theoretical sense of compatability—because imperatives just aren’t the sort of thing that can be true or false. In particular, recommendations of the form above aren’t consistent or inconsistent with the claim that nothing is true.¹¹

Still, the revolutionary fictionalist’s recommendation is usually informed by theoretical considerations, like error theory or nominalism about the topic of interest; and of course he will have to make descriptive claims in order to explain how his particular position differs from other versions of fictionalism and other forms of anti-realism. Not to mention the indicative sentences he uses when engaging in the relevant

¹⁰ See e.g. Rosen (2005).

¹¹ Or if you like: they are vacuously incompatible with everything. In this sense, everything non-indicative is alethic.

discourse. Thus the motivation for and/or practice of revolutionary fictionalism may be essentially alethic, even though its central recommendation isn't. In these respects, the position is just like its hermeneutic counterpart, so we can afford to treat the two together for much of what follows. In fact, some reasons for thinking that fictionalism is an essentially alethic view transcend the prefix/pretense division as well. Let's consider one or two of these first, and then turn to arguments that target prefix and pretense versions of fictionalism individually. Along the way, we'll comment briefly on Rosen and Yablo too.

To begin with, recall that the fictionalist can found his position on nominalism or agnosticism rather than error theory, and thereby avoid making any ascriptions of truth or falsity at the very outset. But the fictionalist about Fs doesn't merely want to have us avoid asserting or believing that there are Fs. His more general aim is to have us avoid asserting or believing anything that *entails* the existence of Fs, and of course entailment is usually taken to be an alethic notion *par excellence*. The fictionalist would probably do best to grant the point, but insist that merely using the notion of entailment to motivate his position doesn't commit him to the existence of truths. On the relevant conception of entailment, to say that certain premises entail a conclusion is just to say that the conclusion is logically guaranteed to be true provided that the premises are true—which is of course compatible with the premises and conclusion not being true.¹² Entailment claims do not themselves entail the unembedded ascription of truth to anything, assuming it's not a logical truth that there are truths.

A different strategy for "tying truth to fictionalism" would spotlight our use of notions like meaning and semantic content in delineating the many faces of fictionalism in the diagram above. As we've seen, prefix fictionalists make claims about what they (and perhaps we) really *mean* when using F-discourse, and pretense fictionalists explain that they (we) take a different sort of attitude toward the same *content* that their realist opponents literally believe. Most philosophers think that at least part of what it is to give the meaning or content of something is to specify the conditions under which it would be true. If the fictionalist shares this popular view, then presumably he will have to admit that some things actually *are* true. After all, he continues to use the sentences whose meanings he gives, so he ought to think that their truth-conditions are sometimes satisfied.¹³

¹² If the reader finds fault with this response, here's another for his consideration. The fictionalist himself doesn't have to *say* that he will refrain from saying or thinking anything that entails the existence of Fs. He just has to refrain from saying or thinking anything that entails the existence of Fs, and he can do this without using the notion of entailment. If the fictionalist wants to say more, there are other responses to the entailment objection available to him. He can say for example that he will refrain from saying or thinking anything within a certain class of claims, a class which happens to contain every sentence that entails the existence of Fs, but which he can specify without using any alethic terms.

¹³ Of course, whatever the fictionalist thinks, if the notion of truth does in fact have a central role to play in the theory of meaning/content, then there will presumably be a sense in which any assertoric use of language brings with it a commitment to the existence of truths. But I take it our objector is looking for a more discriminating sense in which fictionalism might be an essentially alethic position.

This objection fires wide of pretense fictionalists, who *make-believe* the contents of the relevant sentences precisely to avoid any commitment to their truth. It's much less clear, however, whether and to what extent the objection wounds the prefix fictionalist. Granted, Rosen's prefix fictionalist effectively offers truth-conditions for statements in the target class, but we've already seen that his style of fictionalism can be reconceived as a semantic addendum (about modal discourse) to a more straightforward prefix fictionalism (about possible worlds). Insofar as the fictionalism in "Rosen-style fictionalism" is just the second of these two components, adjudicating the case boils down to adjudicating the case of paraphrase-free prefix fictionalism.

Recall in this connection that a charitable reading of simple prefix fictionalism involves distinguishing speaker-meaning from semantic meaning. Giving the semantic meaning of a sentence may well involve specifying its truth-conditions in some objectionable sense, but the usual way to give the speaker-meaning of an utterance is just to offer some clarifying reformulation or indication of subtext or whatever. It hardly seems as though the fictionalist who occasionally makes his prefix explicit is thereby evincing a commitment to model-theoretic semantics.

This consideration is perhaps more conclusive with respect to revolutionary prefix fictionalism than its hermeneutic counterpart. As we've seen, the appeal to speaker-meaning can seem more suspect when the speakers in question are ordinary folk. If the hermeneutic prefix fictionalist is forced to frame his view as a systematic account of the semantic contents of certain swaths of public discourse then he might well find himself saddled with a commitment to truths. It's worth mentioning in this connection that mainstream semantics *might* turn out to be compatible with deflationism about the notion of truth.¹⁴ If so, and that's admittedly a big "if," the fictionalist should in principle be able to engage in the theory of meaning and then paraphrase away his commitment to truths. But these issues are obviously too large to pursue in any detail here. Suffice it to say for present purposes that I'll primarily be interested in revolutionary fictionalism in the next section.

Continuing to target prefix fictionalism specifically, the objector might argue that any plausible analysis or explication of such a prefix will inevitably involve the notion of truth. What exactly does it mean to say, for example, that according to Conan Doyle's stories, Holmes wore plaid? Maybe it just means that, if the stories had been *true*, then Holmes *would* have worn plaid. Alternatively, on a simplified version of David Lewis's well-worn proposal: if the stories had been told as *known fact* (in a world where the audience shares the "common" beliefs of Conan Doyle's actual audience), then Holmes would have worn plaid. The present concern is not, as before, that the prefix fictionalist embroils himself in semantics generally. It's rather that he owes us some account of the meaning of a particular linguistic item, which may well turn out to involve illicit, ineliminable use of a notion like "truth" or "fact."

We've pointed to two different candidate analyses of fictionalization prefixes. The first should concern us less than the second, for the token of "true" in its antecedent

¹⁴ See e.g. Williams (1999).

simply serves a grammatical purpose. Assuming that the term “Conan Doyle’s stories” picks out a string of indicative sentences, we can just take their conjunction as the antecedent in our analysis of this example, without using any alethic notions (besides those used in the stories themselves). Qualms regarding this assumption can be used to motivate Lewis’s theory, which needs to be treated separately. The literature on “truth in fiction” displays a variety of objections to Lewis, one of the more serious of which is his failure to provide a plausible account of the contents of deeply inconsistent fictions.¹⁵ Writers like Gregory Currie and Kendall Walton have developed sophisticated alternatives to Lewis’s view; but we don’t have to survey the metaphysics of aesthetics in order to see that the present objection fails.

In the first place, the revolutionary fictionalist can analyze his prefix however he pleases, since his project is prescriptive rather than descriptive. Even if Lewis’s semantics for ordinary English story prefixes turns out to be correct, the revolutionary fictionalist can simply stipulate that *his* prefix is to be analyzed as per the first suggestion above: as a counterfactual whose antecedent contains at worst a disquotational use of “true.” The hermeneutic prefix fictionalist can’t get off so easily; the interpretation of his prefix is hostage to empirical fortune. But nothing we have said so far suggests that sentences featuring ordinary fictionalization prefixes are alethic in our sense. Suppose for example that, had Conan Doyle’s stories been told as known fact, Holmes would have worn plaid. It doesn’t follow from this that there *are* any facts, or even that there could have been. The point parallels precisely our discussion of the entailment objection above.

But what about pretense fictionalism? Thinking along these same lines, one might worry that the best way to cash out the contrast between belief and make-belief will illicitly involve the notion of truth. Whether or not one’s tempted to say that belief is just the cognitive attitude of “taking as true,” one might naturally think that any plausible account of pretense will involve the idea that pretending does *not* bring with it a commitment to truth. But by now it will be obvious that this is just another illustration of the fact that the explication of a thesis can feature the concept of truth without that thesis entailing that there are truths. Consider the following analogy. Even if turns out that nothing is really morally right or wrong, we might still like to use these concepts in the explanation of psychological phenomena like praise and blame. Thinking of someone *as* having done something morally wrong may well be an essential part of what it is to blame them for what they’ve done.¹⁶ Something similar could be said about the concept of truth and the phenomena of belief and make-belief.

¹⁵ One can remedy this shortcoming by supplementing the standard Lewis/Stalnaker semantics for subjunctive conditionals with a network of impossible worlds. Depending on the details of one’s favorite theory of the nature of worlds, possible and impossible, this amended view may or may not make essential use of the notion of truth.

¹⁶ Of course, if error theory is the right metaethics (and supposing what we’ve said about blame is correct), then we might want to stop blaming people. By the same token, if belief involves “taking as true,” then we might want to stop believing altogether if it turns out that nothing whatsoever is true. Still, the correct account of the psychological difference between a believer and a pretender may involve the notion of truth.

Insofar as Yablo-style fictionalism is an elaboration of pretense fictionalism, the foregoing point applies to it too. But Yablo complicates matters by invoking his distinction between literal and figurative content. If a systematic account of the route from the former to the latter involves the machinery of workaday formal semantics, then there may well be reason to worry that Yablo-style fictionalism is essentially alethic. At the same time, if the literal content of something is determined by its truth-conditions, then presumably its figurative content isn't. Anyway, suffice it to say for present purposes that the considerations in this section provide a strong *prima facie* case for the view that fictionalism—in at least four of the eight main versions described in the previous section—is compatible with nominalism about truths. What will matter most in the next section is just that revolutionary prefix fictionalism in particular is not an essentially alethic position.

18.3 FICTIONALISM ABOUT TRUTH(S)

If some versions of fictionalism are indeed compatible with nominalism about truths, then instances of those versions *about* truths ought to be at least coherent. We might call fictionalism about truths *alethic fictionalism* for short, so long as we're careful not to confuse this use of "alethic" with the notion at play in the pages above. (Alethic fictionalism is not an alethic thesis.) I'll argue presently that fictionalism about truths isn't just an idle option in logical space, but can actually be motivated by reflection on the apparently inevitable "revenge" of semantic paradoxes.¹⁷ With respect to the liar paradox in particular, revenge takes something like the following form.

Any effort to solve a given version of the liar will involve diagnosing some defect in a certain class of sentences. The vocabulary used to make that diagnosis can then be exploited to formulate a new version of the paradox—a sentence that effectively attributes the defect in question to itself. (For instance, if we say that "This sentence is untrue" is defective because it's neither determinately true nor determinately untrue, then what about "This sentence is not determinately true"?) Attempting to treat this "strengthened" liar sentence with the original solution would be tantamount to asserting the strengthened liar. Blocking the inference from that assertion to the claim that the strengthened liar is true would do violence to our ordinary notion of truth. But this ascription of truth is presumably incompatible with the claim that the strengthened liar is defective in the original way. Some new solution will be needed, bringing in its train a new, super-strengthened liar . . .

The so-called inconsistency theory of truth, championed in different forms by Alfred Tarski and Charles Chihara (and more recently, Matti Eklund and Douglas Patterson), purports to explain this tenacity of semantic paradoxes by maintaining that some

¹⁷ Beall (1998) collects recent papers on revenge. See also Sorensen (ch. 25 in this volume).

suitably unrestricted version of the T-schema is indeed “built in” to our very concept of truth. But most inconsistency theorists don’t want to conclude on this basis that contradictions can sometimes be true. One way to avoid this kind of dialetheism would be to argue that the inconsistency theory leads more plausibly to the position that truth itself does not exist. After all, if there were such a property as truth, it would presumably validate the T-schema, which intuitively nothing can. One might then try to found some kind of alethic fictionalism directly on nihilism about truth, or use nominalism about truths to bridge the gap.

There are, however, reasons for the inconsistency theorist to resist this line of reasoning. First of all, if the T-schema is part of the meaning of the truth predicate, then evidently the predicate is meaningful. (One might say for example that there are inconsistent rules governing its use.) And there is something to be said for the pleonastic view that a predicate expresses a property just so long as it’s meaningful. Second, taking a cue from Eklund, one might argue that incoherent predicates just express whatever properties “come closest” to satisfying the constraints imposed by their meanings. Finally, it’s not at all clear whether the notion of property expression is any more coherent than other semantic notions well-known to foster paradox, such as singular reference, application, and satisfaction.¹⁸ The inconsistency theorist may therefore do best not to insist on nihilism about truth. Nominalism or agnosticism about the existence of truths, by contrast, seems to me perfectly coherent and sensible morals to draw from the inconsistency theory.

If indeed our concept of truth is defective, the responsible thing to do would presumably be to revise it, replace it, or just renounce it if all else fails. Much of the technical literature on the paradoxes can be reconceived as offering up candidate revisions to, or replacements for, our ordinary notion of truth. At an admittedly dizzying level of abstraction, one strikingly stable feature of the historical progression of contributions to this literature is a certain trade-off between simplicity and faithfulness. It would seem that the closer a formal theory of truth comes to capturing the intuitive meaning or extension of the predicate, the more complicated it will have to be. Witness for example the leap from Tarski to Kripke. Unfortunately, the more intricate a revision to our ordinary notion of truth, the more difficult it is to internalize. Complexity obviously can’t be a decisive objection against *descriptive* semantic theories; language is a complicated phenomenon. But the degree of complexity exhibited by technically sophisticated responses to the paradoxes does seem to disqualify them as practicable revisionary suggestions. Perhaps a few brave philosophers could adapt their idiolects to reflect Hartry Field’s latest theory of truth, but the rest of us would still be stuck with inconsistency.¹⁹

¹⁸ Of course, this finding would equally undercut the view that the truth predicate *does* express a property, and thus the first and second considerations just cited.

¹⁹ One natural response the revisionist could make to this objection from complexity would involve an appeal to the division of linguistic labor. According to the sociological strand of semantic externalism, certain patterns of deference to expert usage or opinion suffice to ensure that ordinary folk speak with the same meanings and think with the same concepts as the relevant specialists in their linguistic communities. But I doubt philosophers of language will ever enjoy this kind of deference. Moreover,

Now, the retention of our ordinary, incoherent notion of truth doesn't pose any practical problems. The paradoxes rarely arise in everyday conversation; and even when they do, our worldly projects are never compromised. It's not as though we're ever tempted to infer disastrous conclusions from liar-style propositions. At the same time, there's at least something intellectually uncomfortable about continuing to use an expression one knows to be incoherent in the relevant sense. Enter the revolutionary alethic fictionalist, offering a compromise between revision and retention, between theory and practice. Perhaps we can keep using the truth predicate, but use it "lightly" from here on, as a figure of speech or convenient fiction. Even if nothing is true—even if nominalism about truths (or for that matter nihilism about truth) is the right conclusion to draw from the inconsistency theory—there should be no obstacle to our talking and thinking as though many things were.

Incorporating an unrestricted version of the T-schema, the fiction of truth will be inconsistent in much the same way that a novel about a time-traveler who kills his grandfather is inconsistent. Crucially, however, engaging with an inconsistent fiction doesn't involve any commitment to contradictory propositions. Take the alethic prefix fictionalist first. Strictly speaking, all he's committed to is the view that, *according to the fiction of truth*, certain contradictions are true. This doesn't entail a contradiction, nor does it even entail that everything is true according to the fiction of truth. Inconsistency doesn't trivialize stories in the way that it trivializes classical theories. Similarly, the alethic pretense fictionalist needn't assert or believe any contradictions, but merely make-believe the conjunctions of liar sentences and their negations. Granted, this may be difficult to do; perhaps even impossible. Try pretending, by comparison, that two plus two is seven, or that murder is morally obligatory. You can certainly act as though you believe some obviously impossible proposition, but presumably there's more to pretending in the sense of interest to the fictionalist than just this. The version of alethic fictionalism to which I'm most attracted is therefore a revolutionary prefix fictionalism, but other versions deserve to be considered seriously as well.

One observation that might point toward a hermeneutic version of the view is that, pre-theoretically, we tend to think truth consists in some sort of correspondence to reality, and at the same time, tend to marginalize semantic paradoxes. There's arguably something odd about this combination of attitudes. If non-philosophers really believed truth to be a "substantive" property like correspondence, they should presumably think that its nature could be illuminated by some scientific research program—which program we could reasonably expect to reveal the fallacy in liar-style reasoning as arising from certain mistaken assumptions about the nature of truth, reference, and associated phenomena. But this is not a very common outlook. The paradoxes usually strike physicists, psychologists, and the proverbial people in the street as logico-linguistic

I happen to be in that grumpy minority who think the case for externalism was never adequately made. Were we philosophers simply to defer to Field on truth without adapting our own usage of the term, it seems to me we wouldn't thereby share his concept in any psychologically interesting sense. But now is not the time to revisit the reasons for thinking that meanings are in the head.

puzzles of no real scientific import or interest. So perhaps we're just pretending that truth is a real, robust property.

One might object to this hypothesis on the ground that most people don't have much to say about what the property of truth is supposed to be like, beyond some version of the correspondence slogan. But this point can be accommodated by observing that every work of fiction is incomplete or indeterminate in some respects. Just as there's no fact of the matter as to whether or not Hamlet had eggs for breakfast on the day Ophelia drowned, the hermeneutic fictionalist might argue, our practice of alethic pretense doesn't settle whether correspondence is supposed to be a matter of structural isomorphism with states of affairs or rather correlation with mereologically simple facts. No wonder, then, that any resolution to such debates in the metaphysics of truth can seem forever out of reach.

The revolutionary alethic fictionalist, by contrast, can ignore our inchoate correspondence intuitions if he likes, and simply take the content of the fiction of truth to be exhausted by some version of the T-schema. Coupled with nihilism about truth itself, this refinement of the view would have a distinctly deflationist flavor.²⁰ I happen to self-identify as a deflationist, inconsistency theorist, and revolutionary prefix fictionalist about truths. But whether or not to think of alethic fictionalism in general as a kind of deflationism may be a terminological issue in the end. As anyone steeped in the literature can attest, it's hard to say precisely what unifies the various 'isms that fly under the deflationist banner.

Let me end with a summary of what's been accomplished in the pages above. Section 18.1 has hopefully provided a useful taxonomy of fictionalisms and clarified the various relationships between them. Rosen's modal fictionalism was construed as a combination of simple prefix fictionalism about possible worlds and a semantic theory for modal discourse; and Yablo's figuralism was presented as a mixture of pretense fictionalism and (a generalization of) prefix fictionalism. Section 18.2 then advanced the central claim of the chapter: that many of the main varieties of fictionalism on offer in the literature aren't essentially alethic positions in our operationalized sense. Which is to say, the fictionalist gambit is compatible with the view that nothing is true. Fictionalism is undoubtedly a semantic position, trading as it does in notions like meaning and content. Some versions of it may even be essentially alethic in the alternative sense of requiring the notion of truth for their formulation; but the coherence of alethic fictionalism is secured by the availability of versions that don't entail the existence of truths.

So what's truth got to do with fictionalism? Not much. But fictionalism doesn't have its own *Oxford Handbook* yet.²¹

²⁰ See e.g. Woodbridge (2005).

²¹ Many, many thanks to Michael Glanzberg for all his work in putting together this volume, for giving me the opportunity to contribute to it, and especially for his insightful comments on an earlier version of the present chapter. As usual, the mistakes are all mine.

REFERENCES

- Beall, Jc. (1998). *Revenge of the Liar: New Essays on the Paradox*. Oxford: Oxford University Press.
- Brock, S. (1993). "Modal Fictionalism: A Response to Rosen," *Mind* 102: 147–50.
- Burgess, J. and G. Rosen (1997). *A Subject with No Object*. Oxford: Oxford University Press.
- Devitt, M. (1984). *Realism and Truth*. Princeton: Princeton University Press.
- Dummett, M. (1982). "Realism," *Synthese* 52: 55–112.
- Eklund, M. (2002). "Inconsistent Languages," *Philosophy and Phenomenological Research* 64: 251–71.
- Field, H. (1980). *Science without Numbers*. Princeton: Princeton University Press.
- Fraasen, B. C. van (1980). *The Scientific Image*. Oxford: Oxford University Press.
- Kalderon, M. E. (ed.) (2005). *Fictionalism in Metaphysics*. Oxford: Oxford University Press.
- Lewis, D. (1978). "Truth in Fiction," *American Philosophical Quarterly* 15: 37–46.
- Patterson, D. (2009). "Inconsistency Theories of Semantic Paradox," *Philosophy and Phenomenological Research* 79: 387–422.
- Rosen, G. (1990). "Modal Fictionalism," *Mind* 99: 327–54.
- Rosen, G. (2005). "Problems in the History of Fictionalism," in Kalderon (2005), 14–64.
- Stanley, J. (2002). "Hermeneutic Fictionalism," *Midwest Studies in Philosophy* 25: 36–71.
- Williams, M. (1999). "Meaning and Deflationary Truth," *Journal of Philosophy* 96: 545–64.
- Woodbridge, J. (2005). "Truth as a Pretense," in Kalderon (2005), 134–77.
- Yablo, S. (2002). "Go Figure," *Midwest Studies in Philosophy* 25: 72–102.

CHAPTER 19

RELATIVE TRUTH

HERMAN CAPPELEN
AND TORFINN THOMESSEN HUVENES

19.1 INTRODUCTION: TRADITIONAL AND CONTEMPORARY RELATIVISM

THE recent history of the word “relativism” is both interesting and confusing. Throughout its history, the term has been used to denote many philosophical positions, but a fairly radical transition happened at the beginning of the twenty-first century. We will start by describing this change. We will label the earlier use “Traditional Relativism” and the later use “Contemporary Relativism.” Our primary focus here will be on Contemporary Relativism, but first we offer a brief introduction to Traditional Relativism and how it differs from the Contemporary Relativism.

Traditional Relativism: In Plato’s dialogue *Theaetetus*, Socrates discusses Protagoras’ statement, “Man is the measure of all things: of the things which are, that they are, and of the things which are not, that they are not” (*Theaetetus*, 152a). This statement has traditionally been read as expressing a relativist position.¹ In philosophy of science, Paul Feyerabend’s (1975) slogan “anything goes” can be seen as leading to a form of relativism. Thomas Kuhn’s (1962) views about scientific progress have also been described as having relativist implications.²

The way “relativism” is used in moral philosophy is particularly interesting for our purposes. In a much-cited 1975 paper, Gilbert Harman defends a version of moral relativism that he describes as follows:

My moral relativism is a soberly logical thesis—a thesis about logical form, if you like. Just as the judgment that something is large makes sense only in relation to one or another comparison class, so too, I will argue, the judgment that it is wrong of

¹ See e.g. Burnyeat (1976).

² See e.g. Shapere (1964).

someone to do something makes sense only in relation to an agreement or understanding. (Harman 1975: 3)³

It is unclear what, if anything, Protagoras, Feyerabend, Kuhn, and Harman have in common, and we will not try to answer that question here. What we want to highlight is that a development occurred just around the beginning of the twenty-first century. Prior to this development, it would be utterly uncontroversial to apply the term “relativism” to the kind of view Harman describes in the passage above. After this development, a new usage emerged according to which Harman’s view does not count as a form of relativism. The view described by Harman is in effect treated as one of the central opponents of or alternatives to relativism.

Let us start with some observations about Harman’s way of using the term “relativism.” If it is used as Harman uses it in the passage just quoted, then those who think that “A is tall” means the same as a sentence of the form “A is tall for an F” are relativists about tallness.⁴ On that construal, relativism is a plausible position not just about gradable adjectives, but also about expressions like “far away” and “3 PM.” In all these instances, a case can be made that the content of the sentences (e.g. “Sam is tall,” “Sam is far away,” and “It is 3 PM”) include a relation to something (a comparison class, a location, and a time zone). Harman’s claim in the above passage is that moral relativism should be understood in the same way. It is therefore natural to read Harman as making the claim that “It is wrong to F” means the same, roughly, as “It is wrong to F according to agreement A.”⁵

When the term “relativism” is used in Harman’s way, there is one important and salient implication. It seems Nora can assert (1) and Jasmine can assert (2) without disagreement.

- (1) It is wrong to steal.
- (2) It is not wrong to steal.

If the sentence uttered by Nora has the content that it is wrong to steal according to one agreement and the sentence uttered by Jasmine has the content that it is not wrong to steal according to a different agreement, they can both be right and they can agree with each other.

In a recent book, David Velleman is arguably even more explicit than Harman when it comes to using “relativism” in this way. While describing the view that he calls “relativism,” he comments on some of the implications of relativism.

This claim implies that when the Kikuyu say that there isn’t anything wrong with female circumcision and the Mbuti say there is, both may be speaking the truth, because one group is speaking of what’s wrong-for-the-Kikuyu while the other is speaking of what’s wrong-for-the-Mbuti. Of the course, the Kikuyu and the Mbuti

³ See e.g. also Harman and Thomson (1996).

⁴ See e.g. Richard (2004; 2008) for a discussion of comparative adjectives in connection with Contemporary Relativism.

⁵ There are some subtle questions about how to interpret Harman’s views. For instance, Harman emphasizes that he is not making a claim about what people mean in the sense that they intend to make an elliptical claim. See e.g. Harman and Thomson (1996: 4–5).

have a practical disagreement: they disagree over how to treat young women. According to moral relativism, however, there is no proposition whose truth is at issue between them. (Velleman 2013: 46)

He goes on to offer the following elaboration:

What members of the community say, however, is simply that A is wrong, a statement that is normatively valenced. The latter should be interpreted as containing an implicit indexical, as in “wrong-for-us,” the reference of “us” being supplied by the context of utterance, so that the statement expresses the fact that A is wrong for members of that community ... (Velleman 2013: 47)

According to the view that Velleman is describing, if Nora asserts (1) and Jasmine asserts (2), they might very well be expressing compatible propositions. Let us suppose that Nora and Jasmine are members of different communities. In that case, Nora is talking about wrongness-for-members-of-Nora’s community and Jasmine is talking about wrongness-for-members-of-Jasmine’s community. That means that there is no conflict between the propositions expressed by (1) and (2).

From Traditional to Contemporary Relativism: In order to understand the contemporary debate about relativism, it is extremely important to note that, in recent years, a new use of “relativism” has emerged. According to this new usage, the views discussed by Harman and Velleman do not count as forms of relativism. Another term—“contextualism”—was introduced to describe their views. In the ensuing debate, contextualism was typically construed as the central opponent of relativism. So the term “relativism” has had a strange and confusing history indeed.⁶

Our focus in what follows will be on this more recent usage. According to this use of “relativism,” the views discussed by Harman and Velleman do not count as relativist. In section 19.2 we present Contemporary Relativism in more detail and describe some of the different versions of the view. In section 19.3 we look at some applications of Contemporary Relativism. In section 19.4 we talk about the way that Contemporary Relativism is motivated. In section 19.5 we discuss various objections to Contemporary Relativism.

19.2 CONTEMPORARY RELATIVISM: RELATIVE TRUTH

The focus of this chapter is going to be on the kind of Contemporary Relativist views that are exemplified by the works of, among others, Andy Egan (2007; 2010); Max Kölbel

⁶ More generally, the change in usage of “relativism” is an excellent case study in how not to develop theoretical vocabulary. Using a term to first denote a theory and then later using the same term to denote another theory that entails the negation of the first theory is something that an intellectual discipline should try hard to avoid.

(2002); Peter Lasersohn (2005); John MacFarlane (2005b; 2014); Mark Richard (2004; 2008); and Tamina Stephenson (2007). The relativist views in question have for the most part been developed in an effort to make sense of certain areas of discourse, such as discourse about matters of taste. This form of relativism (from now on we will use “relativism” to denote this contemporary version) is often presented as an improvement over more traditional views about the semantics of the relevant natural language expressions. In particular, it is often claimed to have important advantages when compared to so-called “contextualist” views. It is therefore useful to introduce relativism by comparing it to contextualism.

According to a contextualist view, sentences that contain the relevant expressions have different contents or express different propositions in different contexts. This is the standard way to think about sentences that contain indexicals like “I” and “here.” A sentence like (3) expresses different propositions in different contexts depending on who the agent or speaker of the context is.

(3) I am hungry.

However, the idea can be extended to a wide range of expressions. For instance, let us consider a simple contextualist view about predicates of taste, such as “tasty.” The basic idea is that a sentence like (4) can express different propositions in different contexts, depending on the relevant standards of taste.

(4) Haggis is tasty.

In one context, it might be that it is Nora’s standards that are relevant, while in a different context, what is relevant might be Jasmine’s standards. In the former case, the proposition expressed would be the proposition that haggis is tasty relative to Nora’s standards. In the latter case, it would be the proposition that haggis is tasty relative to Jasmine’s standards. Given the traditional use of “relativism,” this view would be labeled “relativist” and given the contemporary use we are focusing on it would be called “contextualist.” Since we are interested in Contemporary Relativism, we will use the term “contextualism” for views like this.

According to contextualists, sentences can have different truth-values relative to different contexts of use in virtue of expressing different propositions. According to a relativist view, there is a more interesting sense in which truth is relative. In order to understand the relativist views that we are going to focus on, it is useful to consider two questions:

- What are the bearers of relative truth and falsity?
- What is truth relative to?

Let us suppose that the relevant bearers of truth and falsity are propositions. Applied to predicates of taste, the relativist idea is that sentences like (4) do not express different

propositions in different contexts. Instead they express propositions that vary in truth-value. For instance, the relevant proposition might be true relative to Nora's standards of taste, but false relative to Jasmine's standards of taste.

However, it is worth emphasizing that it is not enough to say that relativism is the view that propositional truth is relative. For instance, it is a common view that propositions are true or false relative to possible worlds, but Contemporary Relativists are usually seen as making a more controversial claim.⁷ The upshot is that it also matters what truth is relative to. But, as will become clear below, here there is a significant amount of variation between different relativist views.

Since different relativists have proposed different ways of developing and implementing their views, we will look at some of the more prominent proposals in the literature. We are going to focus on the different versions of relativism defended by Lasersohn (2005); MacFarlane (2005b; 2014); and Egan (2007). These views are similar in that they agree that there is an interesting sense in which truth is relative. They allow that it is propositional truth, and not just the truth of sentences, that is relative. In that respect they differ from the contextualist views described above.

But there are also differences between the relativist views. The most salient difference is arguably that relativists have different views about what propositions are true or false relative to. Lasersohn and Egan claim that propositions are true or false relative to individuals. But according to MacFarlane, it is only if propositions are true or false relative to a contexts of assessment that we get a genuine relativist view.⁸ A more subtle difference is that the different authors emphasize different sources of inspiration for their relativist views. For instance, while Lasersohn's implementation of relativism is based on the semantic system developed by David Kaplan (1989), Egan's version of relativism draws inspiration from David Lewis's (1980) account of *de se* attitudes.

19.2.1 Lasersohn: context and circumstance

In order to implement their ideas, some relativists have drawn inspiration from the semantic system developed by Kaplan (1989). According to Kaplan, an expression is associated with two kinds of meaning. The first kind of meaning is the content. It can be represented as a function from circumstances of evaluation to extensions. In the case of a sentence, its extension is a truth-value. The second kind of meaning is the character. It determines the content relative to a context and can be represented as a function from contexts to contents. In the case of indexicals like "I" and "here," the character is a non-constant function. It determines different contents relative to different contexts.

⁷ See e.g. Cappelen and Hawthorne (2009: ch. 1) for further discussion.

⁸ Other relativists have different views on this matter. For instance, according to Kölbel's (2002) version of relativism, propositions are true or false relative to perspectives. It is an open question how deep some of these differences really are.

For present purposes, what is important is that in this system, the truth-value of a sentence can depend on both the context insofar as it plays a role in determining the content, and the circumstance of evaluation with respect to which it is evaluated. In Kaplan's original system, circumstances of evaluation are treated as world-time pairs. That means that the content of a sentence can vary in truth-value across worlds and times. However, relativists have suggested including other parameters alongside the world and the time parameters. Lasersohn argues that this allows relativists to implement their views by making a small adjustment to Kaplan's system:

All we have to do is assign words like *fun* and *tasty* the same content relative to different individuals, but contextually relativize the assignment of truth values to contents, so that the same content may be assigned different truth values relative to different individuals. This allows for the possibility that two utterances express identical semantic content, but with one of them true and the other one false.

This is not at all hard to work out formally, and in fact can be implemented in Kaplan's system with a relatively small adjustment. (Lasersohn 2005: 662)

Lasersohn's proposed adjustment amounts to treating circumstances of evaluation as world-time-individual triples rather than world-time pairs. This makes it possible for the content of a sentence to have different truth-values relative to different individuals or "judges," as he puts it. For instance, if we were to follow Lasersohn and apply this to predicates of taste, the content of (4) could be true relative to Nora and false relative to Jasmine.

(4) Haggis is tasty.

However, it is worth noting that while Lasersohn's specific proposal is to add an individual parameter to the circumstance of evaluation, the basic idea can be generalized. For instance, instead of contents being true or false relative to individuals, they could be true or false relative to standards of taste. Other parameters can also be introduced to handle other expressions. The more general idea is to add an additional parameter to the circumstance of evaluation.

This way of implementing relativism also makes it easy to state the difference between contextualist and relativist views. According to a traditional contextualist theory, the characters of the relevant expressions will be non-constant functions, just like the character of "I" or "now." They will have different contents in different contexts. However, according to the relativist theory, it is not the contents of the sentences that vary, but the truth-values of the sentences.⁹

⁹ While this looks like a straightforward way of implementing the relativist ideas, we have not said much about the role played by Kaplan's (1989) notion of content. This raises certain questions. It is for instance worth taking into account Lewis's (1980) distinction between the semantic value and the propositional content of a sentence. In a compositional semantic theory, the semantic values of complex expressions are determined by the semantic values of its parts and the way they are

19.2.2 MacFarlane: assessment sensitivity

A closely related and prominent version of relativism, developed by MacFarlane (2005b; 2014), is based on the notion of assessment-sensitivity. The idea that sentences are true or false relative to contexts of use should already be familiar. This is the kind of context-dependence we find in cases involving indexicals like “I” and “here.” But, according to MacFarlane, we need not just contexts of use, but also contexts of assessment. A context of assessment is a context in which the use of a sentence is assessed (MacFarlane 2005b: 325). The idea is that the truth-value of a sentence or a proposition can depend, not just on the context of use, but also on the context of assessment. A sentence or proposition that has different truth-values relative to different contexts of assessment is assessment-sensitive.

Assessment-sensitivity allows us to say that a sentence or proposition is true as assessed by one individual, but false as assessed by another. For instance, if Nora were to assert (4), the sentence could be true relative to Nora’s context of assessment, but false relative to Jasmine’s context of assessment.¹⁰

(4) Haggis is tasty.

MacFarlane’s preferred version of relativism differs from a traditional contextualist view in two ways. The first point is that the variation in truth-value is not a matter of different propositions being expressed in different contexts. The second point is that it is the context of assessment, not the context of use, which is relevant.

combined. It is tempting to identify the semantic value of a sentence with its propositional content, with the propositional content playing the role as the objects of speech acts and propositional attitudes. However, Lewis argued that there are good reasons to keep these notions apart. Dummett (1991: 48) draws a similar distinction between ingredient sense and assertoric content. For the purpose of the present discussion, we will assume that relativists are making a claim about propositional content and that we can talk about propositions and contents more or less interchangeably. But insofar as we want to distinguish between the semantic value and the proposition content of a sentence, it is an interesting question whether this claim also goes together with a claim about the semantic value.

¹⁰ The official statement of relativism that is given by MacFarlane (2005b: 328) is that there is at least one assessment-sensitive sentence. According to this approach, what is crucial to relativism is not the idea that propositional truth is relative, but the idea that truth is relative to contexts of assessment. Having said that, his preferred version of relativism still amounts to a view about propositional truth. Following Kaplan (1989), MacFarlane takes propositions to be true or false relative to circumstances of evaluation. That means that the context of assessment can play a role in determining the truth-value of a sentence by playing a role in determining the proposition expressed or by playing a role in determining the relevant circumstance of evaluation. It is the latter role that MacFarlane focuses on. For instance, we can suppose that circumstances of evaluation consist of a world determined by the context of use and a standard of taste determined by the context of assessment. In that case, we can say that the proposition expressed by (4) is true relative to the world of the context of use and Nora’s context of assessment, but false relative to the world of the context of use and Jasmine’s context of assessment. That means that (4) is assessment-sensitive, but only in virtue of expressing an assessment-sensitive proposition. See e.g. also MacFarlane (2014: ch. 3) for relevant discussion.

MacFarlane's distinction between contexts of use and contexts of assessment also allows us to recognize two positions that might not otherwise be salient: content relativism and nonindexical contextualism.

Content Relativism: While MacFarlane focuses on the role that the context of assessment plays in determining the circumstance of evaluation, it is also possible to develop a view according to which the context of assessment plays a role in determining the proposition expressed. Brian Weatherson (2009) calls this view "indexical relativism" and argues that it can be used to give a better theory of indicative conditionals. Herman Cappelen (2008a; 2008b) has explored a similar view that he calls "content relativism." For the purpose of the present discussion, we will ignore these views and focus on relativism as a view about propositional truth.

Nonindexical Contextualism: MacFarlane (2009) also distinguishes between relativism and another view that he calls "nonindexical contextualism." Relativists and nonindexical contextualists agree that propositional truth is relative. The difference comes down to whether the relevant parameters of the circumstance of evaluation are determined by the context of use or the context of assessment. According to relativism, the relevant parameters are determined by the context of assessment. According to nonindexical contextualism, the relevant parameters are determined by the context of use. For instance, the world parameter would be treated along nonindexical contextualist lines rather than relativist lines.

How should we understand the relationship between MacFarlane's version of relativism and the view defended by Lasersohn (2005)? MacFarlane differs from Lasersohn insofar as he makes a point of distinguishing between the parameters of the circumstance of evaluation that are determined by the context of use and the parameters that are determined by the context of assessment. If we look at how Lasersohn (2005: 666) defines truth for sentences in context, it is tempting to think of his view as a version of nonindexical contextualism. However, Lasersohn also makes some comments that indicate that it is not quite as simple as that and he makes a point of insisting that the individual or "judge" parameter is not uniquely determined by objective features of the utterance situation.

If we claim that it is always possible to determine on an objective basis who the judge is, we effectively introduce into our system a level at which truth values are always assigned objectively . . . In order to maintain an authentically subjective assignment of truth values to sentences containing predicates of personal taste, we must allow that the objective facts of the situation of utterance do not uniquely determine a judge. (Lasersohn 2005: 668–9; see also Lasersohn 2009: 363)

In this way, the individual or "judge" parameter differs from the world and time parameters.¹¹ In any case, it is not the purpose of the present discussion to settle what

¹¹ See e.g. Lasersohn (2013) for further relevant discussion.

it takes for a view to count as “genuinely relativist.” The point is to highlight some of the differences and similarities between different versions of relativism.

19.2.3 Egan: centered worlds

Egan (2007; 2010) has developed a version of relativism that draws inspiration from elements of Lewis’s (1979) theory of *de se* attitudes. According to Lewis, the objects of beliefs and desires are not sets of possible worlds, but properties or sets of centered worlds, sets of world-time-individual triples. The basic idea is that if I have the first-person (what Lewis calls “*de se*”) belief that I am hungry, I am self-ascribing the property of being hungry.¹² For present purposes, it will be convenient to assume that it does not matter whether we talk about properties or sets of centered worlds. Insofar as they are meant to play the role of objects of beliefs and desires, let us use Egan’s (2007; 2010) terminology and call them “self-locating contents” and let us talk interchangeably about believing and self-ascribing self-locating contents.

Lewis originally put forward his theory as a solution to problems involving *de se* thought, but his theory has also become an inspiration for relativists like Egan. This kind of proposal is similar to Lasersohn’s (2005) proposal insofar as it posits contents or propositions that vary in truth-value across individuals. Lasersohn (2009: 373) also recognizes the similarity between his relativist theory and Lewis’s (1979) theory of *de se* attitudes. It is nevertheless interesting to see the role that self-locating contents play in Egan’s version of relativism, especially the role they play in his account of assertion.

Whereas Lewis was mainly interested in the contents of mental states, such as beliefs and desires, Egan (2007; 2010) emphasizes the role of self-locating contents as the objects of assertions. The picture of assertion that Egan is working with is more or less that of Robert Stalnaker (1978). The idea is that accepting an assertion requires believing its content. In order to see how this is supposed to work, let us suppose that Nora asserts (4).

(4) Haggis is tasty.

Following Egan (2010), let us further suppose that that the object of her assertion is a self-locating content, something like the property of being disposed to enjoy haggis. In order for Jasmine to accept her assertion, she must then self-ascribe that property.

It is important that Jasmine self-ascribes the property of being disposed to enjoy haggis and that she does not just ascribe this property to Nora. This is an important part

¹² See e.g. Cappelen and Dever (2013) for critical discussion of Lewis’s (1979) theory of *de se* attitudes.

of how the theory handles acceptance and disputes. Egan wants Jasmine to accept Nora's assertion only if she takes herself to be disposed to enjoy haggis. It is not enough that he believes that Nora is disposed to enjoy haggis. We will return to some related issues concerning disagreement in sections 19.4.2 and 19.5.2.

It is also important that the object of the assertion is not the property of being such that Nora is disposed to enjoy haggis. In that case, the truth-value would not vary across different individuals in the same world and we would again fail to predict the desired patterns of acceptance among speakers. This would be more in line with what we should expect from a contextualist theory.

In order to further appreciate this point, it is also useful to look at how Egan wants to treat sentences that involve first-person indexicals like "I." He argues that "the very first place in which one might be inclined to look for self-locating content in natural language—sentences involving first-person indexicals—is not in fact a good place to look" (Egan 2010: 279). In order to see why, let us suppose that Nora asserts (3).

(3) I am hungry.

If the object of her assertion were the property of being hungry, as one might expect, Jasmine would have to self-ascribe the property of being hungry in order to accept her assertion. But that is not right. What she should come to believe is that Nora is hungry, not that she herself is hungry. In order to avoid this unfortunate result, Egan argues that we should accept a standard theory of indexicals like "I" according to which they refer to different individuals in different contexts.¹³

19.3 RELATIVIST HUNTING

GROUND: APPLICATIONS OF RELATIVISM

Contemporary Relativists like Lasersohn, MacFarlane, and Egan are not global relativists. Their relativist views are reserved for restricted domains of discourse. A significant part of the recent literature has focused on two domains: personal taste and epistemic modality. But the relativist project is ongoing and a number of different applications have been proposed and it is likely that more will emerge in the future.

¹³ It is worth noting that according to Egan, while there is a difference between sentences involving predicates of taste and sentences involving first-person indexicals like "I," there is no corresponding difference at the level of thought. The content of my belief that haggis is tasty and the content of my belief that I am hungry both have different truth-values relative to different individuals. This suggests that insofar as there is something special about sentences involving predicates of taste, at least as compared to sentences involving first-person indexicals, we should look for evidence of this at the level of language.

Predicates of Taste: So-called “predicates of personal taste,” or just “predicates of taste,” are among the expressions that have received a lot of attention in the debate about relativism. Several relativists, including Kölbel (2002; 2009); Lasersohn (2005); MacFarlane (2007); Stephenson (2007); and Egan (2010), have argued that predicates of taste ought to be given a relativist treatment.

Epistemic Modals: Epistemic modals are another class of expression that is frequently discussed in connection with relativism. MacFarlane (2011: 144) describes epistemic modals as epistemic uses of modal expressions like “might,” “must,” “possibly,” “probably,” and so forth. Egan, Hawthorne, and Weatherson (2005); Egan (2007); Stephenson (2007); and MacFarlane (2011) have all defended a relativist view about epistemic modals. It is natural to think of epistemic modals as being somehow dependent on a relevant body of information or knowledge. Relativism allows us to capture this by saying that sentences that contain epistemic modals express propositions that have different truth-values relative to different bodies of knowledge or information. According to MacFarlane (2011), this can be implemented by taking sentences like (5) to express assessment-sensitive propositions.

(5) Harry might be in Boston.

For instance, it could be that the proposition that Harry is in Boston is compatible with what Nora knows, whereas Jasmine knows that Harry is not in Boston. That could make it the case that the proposition expressed by (5) is true relative to Nora’s context of assessment, but false relative to Jasmine’s context of assessment.

Other Applications: While predicates of taste and epistemic modals have received a lot of attention, relativism has a number of other potential applications. Here we will only briefly mention some of them. For instance, in addition to being interested in epistemic modals, relativists have also taken an interest in problems involving so-called “deontic modals.” Deontic modals are deontic uses of modals like “may,” “ought,” and “must.” Niko Kolodny and John MacFarlane (2010) suggest that deontic modals, like epistemic modals, are sensitive to an information state and that this should be cashed out in relativist terms.

Relativism has also been proposed as a way of dealing with knowledge ascriptions. Knowledge ascriptions have been the subject of much debate and a number of views have been proposed, including contextualist views.¹⁴ Relativists like Richard (2004; 2008) and MacFarlane (2005a) have argued that relativism is an improvement over the other views in the debate. In particular, relativism avoids some of the problems with contextualist views.

MacFarlane (2003; 2008) has also argued that a relativist view can provide an account of so-called “future contingents,” contingent claims about the open future. For

¹⁴ Contextualism about knowledge ascriptions has been defended by e.g. Cohen (1988; 1999); DeRose (1992; 2009); and Lewis (1996).

instance, let us consider a sentence like “There will be a sea battle tomorrow.” If the future is genuinely “open,” it is unclear what we should say about the truth-value of a sentence like this. MacFarlane argues that relativism allows us to respect the idea that the future is genuinely open without saying that our talk about the open future is systematically confused.

19.4 WHY RELATIVISM? THE CENTRAL MOTIVATIONS

While the different applications of relativism might seem diverse, there are certain themes that are common, even if not universal. In what follows we are going to elaborate on the role that subjectivity and disagreement play in motivating relativism.

19.4.1 Subjectivity

There seems to be an element of subjectivity in many of the areas of discourse for which a relativist treatment has been proposed. This is perhaps most apparent in the case of predicates of taste, but it is also relevant in the case of other areas, such as epistemic modals. Kölbel emphasizes the importance of being able to make sense of non-objective matters:

Arguably, not all natural-language sentences concern objective matters. Many philosophers, and indeed non-philosophers, would deny that it is an objective matter whether a work of art is beautiful. Many would deny that it is an objective matter whether a stew is tasty . . . [My] aim . . . is to show how the assumption of global truth-evaluability *can* be made compatible with the view that not everything is objective. (Kölbel 2002: 19; original emphasis)

MacFarlane also points to subjectivity as a part of the motivation for relativism:

On the one hand, we want to capture the *subjectivity* of the issue—the degree to which their truth seems to depend not just on how things are with the objects they are explicitly about, but on how things are with certain subjects. (MacFarlane 2007: 20; original emphasis)

It is not always clear what it takes for an area of discourse to count as subjective, but when MacFarlane talks about subjectivity, he talks about whether the truths in the domain are also dependent on a subject who is not explicitly mentioned. For instance, in

the case of (4), the idea would be that its truth depends not just on what haggis is like, but also on the relevant subject.

(4) Haggis is tasty.

Relativism seems to be in a good position to capture this kind of subjectivity. A relativist can say that (4) expresses a proposition that has different truth-values relative to different individuals or assessors. This is one way in which the truth-value can depend on a subject that is not explicitly mentioned.

But relativism is not the only way of making sense of subjectivity. This is something that relativist views have in common with contextualist views. According to a contextualist view, a sentence like (4) expresses different propositions in different contexts, depending on the relevant subject or subjects. This is another way in which the truth-value can depend on a subject that is not explicitly mentioned.

More generally, both contextualism and relativism can be seen as benefiting from arguments to the effect that certain sentences do not have a stable truth-value. However, relativism also benefits from arguments that purport to show that there is stability at the level of contents or propositions. In section 19.4.2 we turn to arguments of this sort.

19.4.2 Disagreement

A significant part of the motivation for relativism is based on arguments against rival views. In particular, it has been important to relativists to argue that relativism has advantages over contextualist views. It is perhaps not difficult to get the impression that there is not much to choose between contextualist and relativist views. That is unsurprising given that “relativism” used to be used as a term for both contextualism and what we are now calling “relativism.”

Both contextualism and relativism seem to be in a position to capture the perceived subjectivity of certain areas of discourse, such as discourse about matters of taste. But relativists argue there are reasons to prefer their views to those of their contextualist rivals. A common strategy among relativists is to argue against contextualism on the basis of considerations involving disagreement.¹⁵ In fact, this is something that most Contemporary Relativists have in common. MacFarlane offers the following remarks on contextualism, arguing that while contextualism can explain subjectivity, it has problems with disagreement:

The contextualist can explain . . . why speakers so readily make claims in these domains on the basis of their idiosyncratic tastes, senses of humour, or knowledge. However, by construing these claims as claims about the speaker (or some

¹⁵ This does not mean that considerations involving disagreement are the only considerations that can be used to support a relativist view. See e.g. Kölbel (2009) and Lasersohn (2009) for relevant discussion.

contextually relevant group), the contextualist makes it difficult to make sense of the *disagreement* speakers perceive in these areas of discourse. (MacFarlane 2007: 19; original emphasis)

In order to see why contextualism is supposed to have problems with disagreement, let us again look at predicates of taste and compare a simple relativist view with a simple contextualist view. Let us suppose that Nora sincerely asserts (4) and Jasmine sincerely asserts (6).

- (4) Haggis is tasty.
- (6) Haggis is not tasty.

According to a simple contextualist view, the proposition expressed by (4) in Nora's context is the proposition that haggis tastes good to Nora, while the proposition expressed by (6) in Jasmine's context is the proposition that haggis tastes good to Jasmine. The worry is that this view fails to predict that Nora and Jasmine disagree. If the contextualist view were correct, there would be no obvious conflict involved. The proposition expressed by (6) is not the negation of the proposition expressed by (4). Nora and Jasmine can believe what each of them says. Their respective assertions of (4) and (6) provide no immediate reason to think that they disagree. That, according to relativists, is a fundamental weakness of contextualism.

Similar problems arise in the case of epistemic modals. Let us suppose that Nora sincerely asserts (5) and Jasmine sincerely asserts (7).

- (5) Harry might be in Boston.
- (7) Harry can't be in Boston.

According to a simple contextualist view, the proposition expressed by (5) in Nora's context is the proposition that it is compatible with what Nora knows that Harry is in Boston, while the proposition expressed by (7) in Jasmine's context is the proposition that it is incompatible with what Jasmine knows that Harry is in Boston. The worry is again supposed to be that we fail to predict that they disagree. The proposition expressed by (7) is not the negation of the proposition expressed by (5).

If we adopt a relativist view, we are supposed to do better. For the purpose of this discussion, we can adopt something like Lasnik's (2005) proposal. This allows us to say that the proposition expressed by (4) is true relative to Nora, but false relative to Jasmine and that the proposition expressed by (6) is false relative to Nora, but true relative to Jasmine. In that case, there is a conflict insofar as the propositions in question cannot both be true relative to the same circumstance of evaluation. If Nora believes that haggis is tasty and Jasmine believes that haggis is not tasty, there is a sense in which their beliefs are incompatible. A similar treatment is available in the case of (5) and (7). Whether this is enough to secure disagreement is something we will return to in section 19.5.2, but the

general idea is that relativism has more resources than contextualism when it comes to explaining disagreement.¹⁶

19.4.3 Eavesdroppers

In addition to focusing on straightforward cases of disagreement, relativists have also emphasized the importance of cases involving eavesdropping.¹⁷ These are cases in which a third party who is not a participant in the conversation is assessing or commenting on the relevant assertion. In many ways, these cases are just an extension of the cases of disagreement that we looked at in section 19.4.2. However, according to relativists, eavesdropper cases show that the problems involving disagreement are not just problems for simple versions of contextualism.

It is often observed that there are many cases of disagreement that only present a problem for a very simple version of contextualism. The importance of working with a more flexible and sophisticated contextualist theory has been emphasized by many opponents of relativism, including Glanzberg (2007); Cappelen and Hawthorne (2009: ch. 4); Janice Dowell (2011); and Jonathan Schaffer (2011). It is possible to account for many cases of disagreement by adopting a more sophisticated version of contextualism. For instance, according to a simple contextualist view about predicates of taste, it is only the tastes of the speaker that matter. Similarly, according to a simple contextualist view about epistemic modals, it is only what the speaker knows that is relevant. But more sophisticated versions of contextualism allow more flexibility when it comes to determining whose knowledge or tastes are relevant. This also makes it easier to explain disagreement. For instance, there is more scope for disagreement if the speaker is making a claim about what is tasty or fun for the members of a group and not just about what is tasty or fun for her.

However, eavesdropper cases are supposed to raise problems for this kind of response on behalf of contextualism. For instance, let us consider a case involving epistemic

¹⁶ It is worth emphasizing that these kinds of arguments can take many forms and appeal to different considerations. It is natural to think of disagreement between individuals as a phenomenon at the level of mental states. As Jackson and Pettit put it, “disagreement . . . is a psychological phenomenon. The production of sentences make public our disagreements; it does not create them” (Jackson and Pettit 1998: 251). But even if this is right, there are still relevant linguistic considerations that play an important role in the arguments of the relativists. For instance, Stephenson (2007: 493) emphasizes that she is focusing a notion of disagreement that is tied to the use of expressions like “no” and “nuh-uh.” Egan (2010) focuses on patterns of acceptance and disputes among speakers. We should also distinguish between cases in which the parties are participants in the same conversation and cases in which they are not, as Richard (2004: 218–19; 2008: 93–4) does. For the most part, these subtleties will not play an important role in the present discussion, but we will look at the role played by cases involving eavesdroppers in section 19.4.3. In any case, it is worth keeping these distinctions in mind when one considers the various arguments and responses that are presented in the literature.

¹⁷ See e.g. Egan (2007) and MacFarlane (2011) for a discussion of eavesdroppers in connection with epistemic modals.

modals. Let us suppose that Jasmine is eavesdropping on Nora's conversation and that Nora is unaware of this fact. Furthermore, let us suppose that Jasmine knows that Harry is not in Boston. In that case, if Nora were to assert (5), it makes sense for Jasmine to say "That's false" or "Nora's wrong."

(5) Harry might be in Boston.

As we pointed out above, the simple contextualist view about epistemic modals described in section 19.4.2 has problems with this.¹⁸ What the eavesdropper adds is this: "Let us suppose that we adopt a more flexible version of contextualism, according to which Nora is making a claim about what is compatible with the combined knowledge of the conversational participants." This still would not include potential eavesdroppers. It would still not explain why it makes sense for Jasmine to say that Nora said something false. MacFarlane observes that if contextualists try to solve this problem by expanding the relevant group such that it also includes eavesdroppers, we end up having to interpret speakers as making claims that are too strong:

To sum up: the arguments that motivate a move from the "for all I know" reading of epistemic modals to the "for all we know" reading also motivate extending the scope of "we" to include not just the participants in the conversation but eavesdroppers, no matter how well hidden or how distantly separated in time and space. "It is possible that p" becomes "p is not ruled out by what is known by anyone who will ever consider this claim." (MacFarlane 2011: 152)

MacFarlane goes on to argue that this would make most ordinary uses of epistemic modals irresponsible. Nora can be in a position to assert (5) even if she has no grounds for asserting that it is not ruled out by what is known by anyone who will consider her claim that Harry is in Boston. If this is correct, contextualists are left with a dilemma. They have to choose between explaining the appropriateness of the original assertion and explaining the eavesdropper's response. The relativists argue that they can avoid the dilemma by saying that the proposition expressed by (5) is true relative to Nora, but false relative to Jasmine or something similar along those lines. In that case, it is supposed to be appropriate for Nora to make the assertion and for Jasmine to say that Nora is saying something false.

19.4.4 Faultless disagreement

The claim that relativism has an advantage when it comes to explaining disagreement is sometimes tied to the idea of faultless disagreement. A case of faultless disagreement

¹⁸ According to a simple contextualist view about epistemic modals, the proposition expressed by (5) in Nora's context is the proposition that it is compatible with what Nora knows that Harry is in Boston. But then it would not make sense for Jasmine to say that Nora said something false only on the basis of her knowledge that Harry is not in Boston. In fact, we can stipulate that Jasmine knows that it is compatible with what Nora knows that Harry is in Boston.

is a case of disagreement in which neither party is wrong or making a mistake. In order for the disagreement to count as “faultless” in the relevant sense, it is not enough that neither party is blameworthy or subject to rational criticism. If someone believes something that is not true, she counts as being wrong or having made a mistake even if her belief is based on seemingly strong evidence.

It has already been observed that Nora and Jasmine seem to disagree when Nora sincerely asserts (4) and Jasmine sincerely asserts (6).

- (4) Haggis is tasty.
- (6) Haggis is not tasty.

But it is also not clear that either of them has made any kind of mistake. The problem is making sense of that without being forced to say that they do not disagree. For instance, if the disagreement is a matter of one party believing a proposition and another party believing its negation, it is not clear how we can avoid saying that one of them makes the mistake of believing a proposition that is not true. This is where relativism comes into the picture. The idea is that relativism allows us to say that the proposition that Nora believes is true relative to Nora and that the proposition that Jasmine believes is true relative to Jasmine. In that case, we are supposed to conclude that neither of them has made a mistake.

Kölbel (2002; 2004; 2009) has been a prominent advocate of this line of thought, with Lasersohn (2005: 662) advancing a similar argument. Kölbel argues that relativism is in a unique position when it comes to being able to make sense of faultless disagreement:

There are disagreements without error, or in other words, some propositions are not objective. However, minimal constraints on truth show that if it is true that p , then it is not true that not- p , and if it is true that not- p , then it is not true that p . So if one thinker believes that p and another thinker believes that not- p , one of them makes the mistake of believing a proposition that is not true. The only way to allow faultless disagreement is therefore to relativize truth to perspectives: one disputant’s belief is true in his or her own perspective, and the other disputant’s contradictory belief is true in his or her own perspective. (Kölbel 2002: 100)

The problem of faultless disagreement is supposed to present a dilemma for non-relativists more generally. Unlike the problem discussed in section 19.4.2, it is not a problem that is specifically tied to contextualism. Nora and Jasmine either believe incompatible propositions or they do not. In the former case, the worry is that the disagreement is not faultless unless one takes the relativist route. In the latter case, the worry is that there is no disagreement.

However, not all relativists take faultless disagreement to play an important role in the case for relativism. Some relativists, such as MacFarlane (2005b; 2007), do not explicitly emphasize the possibility of faultless disagreement as a motivation for relativism. In fact, Richard (2008: 132) argues that faultless disagreement remains problematic even

for relativists. One should therefore be careful about assuming that relativists have to be committed to the possibility of faultless disagreement or that faultless disagreement is an essential part of the motivation for relativism.

19.5 PROBLEMS FOR RELATIVISM

Relativism has faced a number of objections. Some critics have focused on the motivation behind specific applications of relativism. For instance, there has been a lot of debate about the role that disagreement plays in motivating relativist views. But there are also objections that purport to show that there is something wrong with relativism on a more general level. The worry is that relativism is somehow incoherent or that the view cannot be properly stated. The plan below is to start by looking at some objections of the latter sort before we take on the debate about disagreement.

19.5.1 General problems

One worry is that relativism is somehow self-refuting. For instance, if someone were to claim that there are no absolute truths in the sense that all propositions are true relative to some individuals and false relative to others, that would entail that the proposition that there are no absolute truths is also not an absolute truth. It would be true relative to some individuals and false relative to others. This is supposed to show that there is something wrong with the relativist position. More needs to be said before something like this can be turned into a genuine argument. Kölbel (2011) discusses attempts to do that. But for present purposes it is important to remember that the kind of Contemporary Relativist views that we are interested in have not been put forward as global views. Contemporary Relativists have instead been making local claims about, say, predicates of taste or epistemic modals. These relativist views are unlikely to be susceptible to straightforward self-refutation worries of this sort.¹⁹

Having said that, there are other general worries that are still relevant. One set of issues concerns assertion and belief. Relativists have recognized the importance of explaining how assertion works from a relativist perspective. For instance, if assertions aim at truth, what does that amount to if truth is relative? MacFarlane (2005b) and Egan (2007) are among those who have attempted to meet versions of this challenge. MacFarlane attempts to answer this challenge by developing an account of assertion in terms of the speaker's commitments whereas Egan attempts to reconcile relativism with Stalnaker's (1978) picture of assertion and communication.

¹⁹ See e.g. Wright (2008) and MacFarlane (2014: ch. 2) for relevant discussion.

Similar issues arise in the case of belief. Aaron Zimmerman (2007) argues that relativism has problems when it comes to making sense of believing propositions that are true or false relative to other parameters than worlds. He argues that it is impossible to accept relativism and believe the relativist propositions without being irrational. For instance, if Nora recognizes that the proposition that haggis is tasty is true relative to her standards, but false relative to other people's standards, she will not believe that proposition. She will only believe the proposition that haggis tastes good relative to her standards. To the extent that these considerations involving assertion and belief present a problem for relativism, they present a problem for both global and local versions of relativism.

Another worry that has been raised by Herman Cappelen and John Hawthorne (2009) is that a relativist might have a hard time distinguishing her position from the position of a realist who does not accept either contextualism or relativism.²⁰ On the one hand, relativists have access to a monadic truth predicate "is true" that obeys standard disquotational principles. This allows them to make statements such as (8).

- (8) Nora believes that haggis isn't tasty. Haggis is tasty. Therefore, what Nora believes is false.

On the other hand, Cappelen and Hawthorne argue that realists can also make sense of constructions such as "true by the standards of so-and-so." Simplifying somewhat, it becomes hard to distinguish the way that relativists talk from the way that realists talk. Again, insofar as this is a problem, it is not just a problem for global versions of relativism.

19.5.2 Too little disagreement

Insofar as disagreement plays an important role in the motivation for relativism, it should not come as a surprise that this is a topic that has attracted a lot of controversy. Some of the critics have raised questions about whether relativist views deliver the right predictions about disagreement. Other critics have argued that relativists have exaggerated the problems that contextualist views face when it comes to explaining disagreement. In what follows, we will look at some examples of both kinds of criticism.

The idea that relativism allows us to make sense of faultless disagreement has proven to be particularly controversial. The possibility of faultless disagreement has even been described as absurd by some commentators, including Jason Stanley (2005: 141) and Michael Glanzberg (2007: 16). Different arguments have been put forward which purport

²⁰ Another interesting question is what distinguishes relativism from sophisticated forms of expressivism and non-factualism like those defended by, among others, Gibbard (1990; 2003) and Yalcin (2007; 2011). For relevant discussion see e.g. Kölbel (2002: 110–15).

to show that relativism cannot deliver faultless disagreement.²¹ For instance, Richard (2008: 132) and Cappelen and Hawthorne (2009: 131) argue that if relativists want to make sense of faultless disagreement, they have to accept the truth of sentences like (9) and they take this to be a problematic consequence.

(9) There is no fault whatsoever in speaking falsely.

However, the criticism has also been directed at the more general claims that relativists have made about disagreement. It is important to the relativists that they are in a position to explain disagreement in cases in which contextualists supposedly fail to do so. For instance, they claim to be in a position to say that Nora and Jasmine disagree when Nora sincerely asserts (4) and Jasmine sincerely asserts (6).

(4) Haggis is tasty.

(6) Haggis is not tasty.

According to the simple relativist story that we considered in section 19.4.2, there is a proposition, the proposition that haggis is tasty, such that Nora believes that proposition and Jasmine believes its negation. That proposition is true relative to Nora, but false relative to Jasmine.

A worry is that the simple relativist story presupposes that it is sufficient for disagreement that there is a proposition such that one party believes that proposition and the other party believes its negation. But if we look at other views according to which propositional truth is relative, this assumption starts to look less plausible. For instance, MacFarlane (2007) observes that tensed propositions, propositions that are true or false relative to times, seem to present a problem for this sufficient condition for disagreement:

Consider, for example, tensed propositions, which have truth values relative to world/time pairs. One such proposition is the proposition that Joe is sitting . . . If you assert this proposition at 2 PM and I deny it at 3 PM, we have not in any real sense disagreed. Your assertion concerned Joe's position at 2 PM, while my denial concerned his position at 3 PM. So accepting and rejecting the same proposition cannot be sufficient for disagreement. (MacFarlane 2007: 22)

MacFarlane goes on to argue that a similar point can be made regarding propositions that are true or false relative to worlds. However, he thinks that his relativist view can provide a solution to the problem since it allows him to distinguish between parameters that are determined by the context of use and parameters that are determined by the context of assessment. While the world and time parameters fall into the first category, the parameters

²¹ See e.g. Stojanovic (2007); Moruzzi (2008); Rosenkranz (2008); and Moltmann (2010) for relevant discussion.

that relativists are interested in fall into the second category. But even if MacFarlane (2007) takes these considerations to provide support for the kind of relativist view that he favors, it is safe to say that not everyone has been convinced. Some critics, including James Dreier (2009) and Ragnar Francén (2010), have questioned whether MacFarlane has succeeded in solving the problem. This remains a point of controversy in the debate.

19.5.3 Too much disagreement

So far we have considered the worry that relativists fail to live up their promises when it comes to predicting enough disagreement. However, there is also a worry that relativist views predict too much disagreement. Cappelen and Hawthorne (2009: 122–4) have pointed out that there are cases involving predicates of taste in which we do not find disagreement. They invite us to consider a scenario in which a child asserts (10) on the grounds that she is going to a music camp, while her parents assert (11) on the grounds that they have to work overtime during the summer.

(10) The summer is going to be fun.

(11) The summer isn't going to be fun.

In this case, it would be strange to say that the child and her parents disagree. But Cappelen and Hawthorne argue that this is what relativists are committed to. For instance, if we look at how relativists treat (4) and (6), it is not clear why (10) and (11) would be different.

(4) Haggis is tasty.

(6) Haggis is not tasty.

In order to avoid this problem, relativists have to find a way of distinguishing between these cases. Lasersohn (2011: 436–7) has for instance proposed that there might be multiple dimensions of context-dependence at work. In particular, he suggests that the summer can be construed as a complex event and that the parties are talking about different parts of it. However, there is a danger that moves like this will make the overall relativist story more complicated.

The worry that relativist views predict too much disagreement also applies to cases involving eavesdroppers. Kai von Fintel and Anthony Gillies (2008: 91) use a case involving epistemic modals embedded under disjunction to illustrate this point. Let us suppose that Bond has planted misleading evidence, hoping that Blofeld will find it and come to believe that Bond is in Zürich. In fact, Bond is in London with Leiter and they are listening to Blofeld through a bug that Bond planted along with the misleading evidence. Blofeld finds the misleading evidence, but he remains suspicious and only asserts (12).

(12) Bond might be in Zürich or he might be in London.

In this case, it looks like it would be inappropriate for Leiter to say “That’s false” or “Blofeld’s wrong” in response to what Blofeld said. But von Fintel and Gillies argue that the relativists cannot predict this. They argue that (12) entails (13) and (14).

(13) Bond might be in Zürich.

(14) Bond might be in London.

If that is correct, the relativists have a problem. The problem is that the proposition expressed by (13) is false relative to Leiter and his context of assessment. That means that relativists have to explain why cases like (12) should be treated differently from the cases discussed in sections 19.4.2 and 19.4.3. Otherwise, it looks like the relativists are predicting too much disagreement.

19.5.4 Contextualism and disagreement

While some critics have focused on the relativist story about disagreement, others have argued that disagreement presents a less serious problem for contextualist views than relativists have argued. If that is correct, a significant part of the motivation for relativism is lost.

An interesting view that has emerged from the debate is the so-called “cloud of contexts” view developed by von Fintel and Gillies (2008; 2011). They primarily discuss this view in connection with epistemic modals, but there is no obvious reason why a similar strategy could not work in other cases as well. When a speaker uses a sentence containing an epistemic modal like “might,” there is a question of whose knowledge or information counts as relevant. According to von Fintel and Gillies, the context does not fully determine whose knowledge counts as relevant. In fact, there are multiple admissible contexts.

There is no such thing as “the context,” only the contexts admissible or compatible with the facts as they are. The context of the conversation really does not provide a determinate resolution and we propose to model this by saying that there is a cloud of contexts at the given point of the conversation. (von Fintel and Gillies 2011: 118–19)

As a result of there being multiple admissible contexts, multiple propositions are “put into play,” as von Fintel and Gillies put it. In order for a speaker to be in a position to make an assertion, she does not have to be in a position to assert all of the propositions that are put into play. She only needs to be in a position to assert one of them. However, the hearer can respond to a stronger proposition, depending on her epistemic position and what counts as a cooperative contribution to the conversation.

This is meant to address at least some of the problems with contextualist views. For instance, a speaker might use a sentence containing an epistemic modal on the basis of being in position to assert one of the weaker propositions that is put into play, such as the proposition that the relevant proposition is compatible with her knowledge.

However, an eavesdropper might be responding to one of the stronger propositions that has been put into play, even if that is a proposition that the speaker would not be in a position to assert.

More generally, it is worth recognizing that what is at issue does not have to be the propositions expressed by the relevant sentences.²² In his discussion of contextualism about aesthetic predicates, Tim Sundell (2011) argues that many cases of disagreement are about the selection and appropriateness of the relevant aesthetic standards. Torfinn Huvenes (2012) argues that cases of disagreement involving predicates of taste can be understood in terms of conflicting non-doxastic attitudes—attitudes other than beliefs, such as desires or preferences.²³ If Nora likes haggis and Jasmine dislikes haggis, the difference in attitudes constitutes a conflict and is sufficient for them to disagree. Huvenes (2014) argues that this is also a way of making sense of faultless disagreement. If the parties disagree in virtue of having conflicting desires or preferences, there is little or no pressure to think that one of them is somehow wrong or has made a mistake. This way of making sense of faultless disagreement also avoids many of the problems facing a relativist account of it.

There is no conflict between these responses. Some responses might work better in some cases, while other responses might work better in other cases. It also matters what kinds of expression we are talking about. For instance, it makes sense to try to explain cases of disagreement involving predicates of taste in terms of the non-doxastic attitudes of the parties. But it is arguably more difficult to see how this is going to work in the case of knowledge ascriptions. In any case, the important point is that contextualists have more resources when it comes to explaining disagreement than is sometimes recognized. If it turns out that relativism is not in a better position than contextualism when it comes to explaining disagreement, that would undermine a significant part of the motivation for relativism.

19.6 CONCLUSION

MacFarlane, in the preface to his book *Assessment Sensitivity*, says:

Analytic philosophers are now considerably more open to relativism about truth than they were when I began this project. My initial aim was merely to place relativist views on the table as real options. Many of those who initially accused these views of incoherence have come around to regarding them as merely empirically false. (MacFarlane 2014: vii)

²² See e.g. von Fintel and Gillies (2008) for a discussion about relativism in regard to epistemic modals, in which they emphasize that it does not have to be the propositions expressed by the relevant sentences that are at issue. For instance, they point out that when a speaker asserts, for example, “Harry might be in Boston,” the reply might be targeting the proposition that Harry is in Boston, the so-called “prejacent.”

²³ See e.g. Huvenes (2015) for a related treatment of epistemic modals in terms of different credences.

MacFarlane is right that the form of relativism we have been concerned with in this chapter is now, for the most part, considered an empirical hypothesis about the workings of natural language. It has strengths and weaknesses, some of them sketched above. No matter how one thinks those will balance out at the end of the day, relativism has emerged as an important and influential framework for thinking about truth and natural language.

We want to end with a reminder and a vague conjecture. The reminder is this: for those who use “relativism” in the way that Harman and Velleman use the term, the debate we have described might seem strange. After all, they would apply “relativism” to both contextualist and relativist views. The vague conjecture is this: perhaps one way to see continuity between the two senses of “relativism” involves a joint concern with hidden (or *unobvious*) relativity. On the one hand, if there is a hidden variable associated with the expression—e.g. an unpronounced “for *x*”—we have one form of hidden (unobvious) relativity. On the other hand, if the truth-value of the proposition expressed by a sentence containing that expression is relative in one of the ways we described in section 19.2, we have another form of hidden (unobvious) relativity. So understood, the Traditional and Contemporary Relativists are concerned with a joint phenomenon and can be seen as giving different ways of characterizing it.

REFERENCES

- Burnyeat, M. (1976). “Protagoras and Self-Refutation in Plato’s *Theaetetus*,” *The Philosophical Review* 85: 172–95.
- Cappelen, H. (2008a). “Content Relativism and Semantic Blindness,” in M. García-Carpintero and M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press, 265–86.
- Cappelen, H. (2008b). “The Creative Interpreter,” *Philosophical Perspectives* 22: 23–46.
- Cappelen, H. and J. Dever (2013). *The Inessential Indexical*. Oxford: Oxford University Press.
- Cappelen, H. and J. Hawthorne (2009). *Relativism and Monadic Truth*. Oxford: Oxford University Press.
- Cohen, S. (1988). “How to be a Fallibilist,” *Philosophical Perspectives* 2: 91–123.
- Cohen, S. (1999). “Contextualism, Skepticism, and the Structure of Reasons,” *Philosophical Perspectives* 13: 57–89.
- DeRose, K. (1992). “Contextualism and Knowledge Attributions,” *Philosophy and Phenomenological Research* 52: 913–29.
- DeRose, K. (2009). *The Case for Contextualism*. Oxford: Oxford University Press.
- Dowell, J. (2011). “A Flexible Contextualist Account of Epistemic Modals,” *Philosopher’s Imprint* 11: 1–25.
- Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Egan, A. (2007). “Epistemic Modals, Relativism and Assertion,” *Philosophical Studies* 133: 1–22.
- Egan, A. (2010). “Disputing about Taste,” in R. Feldman and T. Warfield (eds.), *Disagreement*. Oxford: Oxford University Press, 247–86.
- Egan, A., J. Hawthorne and B. Weatherson (2005). “Epistemic Modals in Context,” in G. Preyer and G. Peter (eds.), *Contextualism in Philosophy*. Oxford: Oxford University Press, 131–68.

- Feyerabend, P. (1975). *Against Method*. London: Verso.
- Fintel, K. von and A. Gillies (2008). "CIA Leaks," *The Philosophical Review* 117: 77–98.
- Fintel, K. von and A. Gillies (2011). "‘Might’ Made Right," in A. Egan and B. Weatherston (eds.), *Epistemic Modality*. Oxford: Oxford University Press, 108–30.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Oxford: Oxford University Press.
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Glanzberg, M. (2007). "Context, Content, and Relativism," *Philosophical Studies* 136: 1–29.
- Harman, G. (1975). "Moral Relativism Defended," *The Philosophical Review* 84: 3–22.
- Harman, G. and J. Thomson (1996). *Moral Relativism and Moral Objectivity*. Cambridge, MA: Blackwell.
- Huvenes, T. (2012). "Varieties of Disagreement and Predicates of Taste," *Australasian Journal of Philosophy* 90: 167–81.
- Huvenes, T. (2014). "Disagreement without Error," *Erkenntnis* (special issue on disagreement, ed. Daniel Cohnitz and Teresa Marques) 79: 143–54.
- Huvenes, T. (2015). "Epistemic Modals and Credal Disagreement," *Philosophical Studies* 172: 987–1011.
- Jackson, F. and P. Pettit (1998). "A Problem for Expressivism," *Analysis* 58: 239–51.
- Kaplan, D. (1989). "Demonstratives," in J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*. Oxford: Oxford University Press, 481–563.
- Kölbel, M. (2002). *Truth Without Objectivity*. London: Routledge.
- Kölbel, M. (2004). "Faultless Disagreement," *Proceedings of the Aristotelian Society* 104: 53–73.
- Kölbel, M. (2009). "The Evidence for Relativism," *Synthese* 166: 375–95.
- Kölbel, M. (2011). "Global Relativism and Self-Refutation," in S. Hales (ed.), *A Companion to Relativism*. Oxford: Blackwell, 11–30.
- Kolodny, N. and J. MacFarlane (2010). "Ifs and Oughts," *Journal of Philosophy* 107: 115–43.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: Chicago University Press.
- Lasersohn, P. (2005). "Context Dependence, Disagreement, and Predicates of Personal Taste," *Linguistics and Philosophy* 28: 643–86.
- Lasersohn, P. (2009). "Relative Truth, Speaker Commitment, and Control of Implicit Arguments," *Synthese* 166: 359–74.
- Lasersohn, P. (2011). "Context, Relevant Parts and (Lack of) Disagreement over Taste," *Philosophical Studies* 156: 433–9.
- Lasersohn, P. (2013). "Non-World Indices and Assessment-Sensitivity," *Inquiry* 56: 122–48.
- Lewis, D. (1979). "Attitudes De Dicto and De Se," *The Philosophical Review* 88: 513–43.
- Lewis, D. (1980). "Index, Context, and Content," in S. Kanger and S. Öhman (eds.), *Philosophy and Grammar*. Dordrecht: Reidel, 79–100.
- Lewis, D. (1996). "Elusive Knowledge," *Australasian Journal of Philosophy* 74: 549–67.
- MacFarlane, J. (2003). "Future Contingents and Relative Truth," *Philosophical Quarterly* 53: 321–36.
- MacFarlane, J. (2005a). "The Assessment Sensitivity of Knowledge Attributions," in T. S. Gendler and J. Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 1*. Oxford: Oxford University Press, 197–233.
- MacFarlane, J. (2005b). "Making Sense of Relative Truth," *Proceedings of the Aristotelian Society* 104: 53–73.
- MacFarlane, J. (2007). "Relativism and Disagreement," *Philosophical Studies* 132: 17–32.
- MacFarlane, J. (2008). "Truth in the Garden of Forking Paths," in M. García-Carpintero and M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press, 81–102.

- MacFarlane, J. (2009). "Nonindexical Contextualism," *Synthese* 166: 231–50.
- MacFarlane, J. (2011). "Epistemic Modals are Assessment-Sensitive," in A. Egan and B. Weatherson (eds.), *Epistemic Modality*. Oxford: Oxford University Press, 144–78.
- MacFarlane, J. (2014). *Assessment Sensitivity*. Oxford: Oxford University Press.
- Moltmann, F. (2010). "Relative Truth and the First Person," *Philosophical Studies* 150: 187–220.
- Moruzzi, S. (2008). "Assertion, Belief and Disagreement," in M. García-Carpintero and M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press, 207–24.
- Richard, M. (2004). "Contextualism and Relativism," *Philosophical Studies* 119: 215–42.
- Richard, M. (2008). *When Truth Gives Out*. Oxford: Oxford University Press.
- Rosenkranz, S. (2008). "Frege, Relativism and Faultless Disagreement," in M. García-Carpintero and M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press, 225–37.
- Schaffer, J. (2011). "Perspective in Taste Predicates and Epistemic Modals," in A. Egan and B. Weatherson (eds.), *Epistemic Modality*. Oxford: Oxford University Press, 179–226.
- Shapere, D. (1964). "The Structure of Scientific Revolutions," *The Philosophical Review* 73: 383–94.
- Stalnaker, R. (1978). "Assertion," in P. Cole (ed.), *Syntax and Semantics, vol. 9: Pragmatics*. New York: Academic Press, 315–32.
- Stanley, J. (2005). *Knowledge and Practical Interests*. Oxford: Oxford University Press.
- Stephenson, T. (2007). "Judge Dependence, Epistemic Modals, and Predicates of Personal Taste," *Linguistics and Philosophy* 30: 487–525.
- Stojanovic, I. (2007). "Talking about Taste," *Linguistics and Philosophy* 30: 691–706.
- Sundell, T. (2011). "Disagreements about Taste," *Philosophical Studies* 155: 267–88.
- Velleman, D. (2013). *Foundations for Moral Relativism*. Cambridge: Open Book Publishers.
- Weatherson, B. (2009). "Conditionals and Indexical Relativism," *Synthese* 166: 333–57.
- Wright, C. (2008). "Relativism about Truth Itself: Haphazard Thoughts about the Very Idea," in M. García-Carpintero and M. Kölbel (eds.), *Relative Truth*. Oxford: Oxford University Press, 157–85.
- Yalcin, S. (2007). "Epistemic Modals," *Mind* 116: 983–1026.
- Yalcin, S. (2011). "Nonfactualism about Epistemic Modality," in A. Egan and B. Weatherson (eds.), *Epistemic Modality*. Oxford: Oxford University Press, 296–332.
- Zimmerman, A. (2007). "Against Relativism," *Philosophical Studies* 133: 313–48.

CHAPTER 20

TRUTH PLURALISM

NIKOLAJ J. L. L. PEDERSEN
AND MICHAEL P. LYNCH

20.1 INTRODUCTION

MONISM about X is the view that there is precisely one way of being X . Pluralism about X is the view that there are several ways of being X . Nihilism about X is the view that there is no such thing as being X . Monists and pluralists disagree over how many ways there are of being X , but they agree that the nihilist is wrong in denying that there is no such thing as being X . Turn now to truth. Monism about truth is the view that there is precisely one way of being true. Nihilism about truth is the view that there is no such thing as being true. And pluralism about truth is the view that there are several ways of being true. For the pluralist, truth is Many; in particular, different ways of being true apply to different domains of discourse. The way in which propositions about physics are true could differ from the way in which the propositions of morality are true.

This chapter provides an introduction to truth pluralism. In section 20.2 we present three different motivations for being a pluralist about truth. Section 20.3 presents a number of challenges and objections to pluralism. Section 20.4 offers a brief survey of different incarnations of truth pluralism. In sections 20.5 and 20.6 we discuss whether different kinds of pluralist have the resources adequately to address the various challenges and objections. Section 20.7 briefly discusses what connections, if any, pluralism about truth might bear to other kinds of pluralism.

20.2 MOTIVATIONS

The basic pluralist idea is that truth is Many. What reasons are there for buying into this idea rather than the basic monist idea that truth is One? We present three reasons drawn from the literature.¹

A principal reason for adopting truth pluralism is that the view provides a framework for understanding the intuitive appeal of, respectively, realism and anti-realism with regard to different domains. The intuitive appeal stems in part from the observation that both traditional realist accounts of truth, such as the correspondence theory, and traditional anti-realist accounts, such as the coherence theory, face a similar pattern of failure. Theories that seem plausible in some domains fail to seem as plausible in others.²

This thought can be fleshed out as follows. Suppose we take realism about a given domain D_i to amount to the thesis that the Principle of Bivalence is characteristic of D_i -discourse. The Principle of Bivalence is a semantic determinacy thesis. Any D_i -proposition is either true or false. Anti-realism about a domain amounts to the thesis that the Principle of Bivalence does not hold in general for the domain. Realism and anti-realism thus understood seem intuitively appealing for, respectively, discourse about macroscopic physical objects and moral discourse. It seems intuitively right that propositions concerning macroscopic physical objects are either true or false—even if they pertain to matters beyond our ken. \langle There are mountains \rangle is true or false, and we can ascertain which of the two it is. It would likewise seem that \langle No star is dying in the universe right now \rangle is true or false even if we are not in a position to know which one, perhaps even as a matter of principle. (Angle brackets are used to indicate propositions.) If we turn to moral discourse, appearances are different. It is not clear that all moral propositions have a determinate semantic status. To illustrate consider the following moral dilemma: during the Second World War, Sophie, a mother of two, is forced by the Nazis to choose which of her two children will die and which will live. Let “Sophie’s Choice” denote the following proposition:

(SC) Letting the older child live and the younger child die is the morally right decision.

Intuitively (SC) does not have a determinate semantic status. Thus, intuitively, the Principle of Bivalence does not seem to hold in full generality for moral discourse.³

¹ As suggested by the specific formulation of the question at hand, the three reasons to be given engage only with the debate between monists and pluralists. We leave nihilism out of the picture for present purposes.

² Lynch (2004: 385) introduces the label the “Scope Problem” for this observation, and this label has become prevalent in the literature on truth pluralism. Sher (1998: 134–5) uses the label “problem of the common denominator” to refer to the same problem.

³ For an example like this, see Lynch (2009: 98).

Pluralists claim to be able to accommodate the intuitive appeal of realism and anti-realism with respect to different domains of discourse. To support this claim let us consider a version of pluralism according to which correspondence and so-called superassertibility govern discourse about, respectively, macroscopic physical objects and morals. A proposition is superassertible if and only if it is warranted in some state of information I_j and remains warranted in any state of information I_k that extends I_j .⁴

Let us refer to discourse about macroscopic physical objects and morals respectively as “ D_{MPO} ” and “ D_M .” Correspondence and superassertibility relate to truth and falsity in D_{MPO} and D_M as follows:

- (COR) For any D_{MPO} -proposition p , p is true if and only if p corresponds with fact, and p is false if and only if p does not correspond with fact.
- (SUP) For any D_M -proposition p , p is true if and only if p is superassertible, and p is false if and only if $\sim p$ is superassertible.

Falsity within D_{MPO} is the complement of truth, because non-correspondence is the complement of correspondence, and because truth within D_{MPO} relates bi-conditionally to correspondence and falsity to non-correspondence. Thus, if $\langle \text{No star is dying in the universe right now} \rangle$ corresponds, the proposition is true. In all other cases—i.e. if it fails to correspond—it falls into the complement of correspondence—i.e. non-correspondence. But, by (COR), this means that the proposition is false. Combining the two observations just made we have that $\langle \text{No star is dying in the universe right now} \rangle$ is true or false.

According to (SUP), truth is bi-conditionally related to superassertibility. In this respect (COR) and (SUP) are similar. However, they are crucially different when it comes to falsity. According to (SUP) falsity is *not* bi-conditionally related to the complement of superassertibility: it is *not* the case that p is false just in case p fails to be superassertible. Thus, unlike in the case of (COR), falsity is not the absence of truth. Instead, according to (SUP), the falsity of p is tied to the superassertibility of $\sim p$. Given the characterization of superassertibility, we thus see that truth and falsity are alike in that they both require a

⁴ Wright (1992: 48). The extendability relation induces a partial order, i.e. it is reflexive, transitive, and antisymmetric:

- *Reflexivity*: for any state of information I_i , I_i extends I_i .
- *Transitivity*: for any states of information I_i , I_j , and I_k , if I_j extends I_i and I_k extends I_j , then I_k extends I_i .
- *Antisymmetry*: for any states of information I_i and I_j , if I_i extends I_j and I_j extends I_i , then $I_i = I_j$.

Wright intends superassertibility to generalize the notion of truth familiar from mathematical intuitionism (1993: 414). The idea that superassertibility applies within the moral domain is spelled out in some detail in Wright (2003b). Wright (1992) takes correspondence to be the most plausible candidate for discourse about macroscopic physical objects. Lynch (2009) also falls in this category. Lynch (2009: ch. 8) suggests that what he calls “concordance” is a plausible candidate for moral discourse. Concordance is different from superassertibility, but like it in crucial respects.

state of information that warrants assertibility. This is why superassertibility does not in general underwrite Bivalence. For some proposition it may be that there is no state that supports it, and also, it may be that there is no state that supports its negation either. To see this return to (SC), (Sophie's Choice). Arguably there is no state of information in which (SC) is warrantably assertible. Hence, (SC) is not superassertible, and so, (SC) is not true. This does not imply that (SC) is false. In order for (SC) to be false \sim (SC) must be superassertible. However, as with (SC), arguably there is no state of information that warrants the assertion of \sim (SC). By considering (SC) we thus see that the Principle of Bivalence does not hold in general for moral discourse.

By bringing together the preceding considerations concerning correspondence and superassertibility we see how the pluralist framework makes it possible to accommodate the intuitive appeal of, respectively, realism and anti-realism with regard to different domains.⁵

A second motivation for truth pluralism is what we might refer to as the "Game Analogy Argument." The Game Analogy Argument draws an analogy between winning and truth. The pluralist observes that there is variation in terms of what winning amounts to across different games. In soccer a team wins if it scores more goals than the other team, while in chess a player wins by checkmating the opponent's king. It is easy to multiply examples. When thinking about games and winning we see that there is a range of properties (scoring more goals than the opponent, checkmating the opponent's king, etc.) that constitute winning for different games. It would thus seem that a form of pluralism about winning is right. What winning amounts to depends on what game is being played, and in this sense there are several ways of winning. According to the pluralist, truth is analogous to winning. There is variation in terms of what truth amounts to across different domains of discourse. For discourse about macroscopic physical objects, a proposition is true if it corresponds with reality, while for moral discourse, a proposition is true if it is superassertible. Other properties—including coherence—might be relevant to other kinds of discourse. There is thus a range of properties (correspondence, superassertibility, coherence, etc.) that constitute truth for different domains of discourse. The truth case is relevantly similar to the case of winning. Hence, if pluralism about winning is right, so too is pluralism about truth. What truth amounts to depends on the domain of discourse, and in this sense there are several ways of being true.⁶

⁵ See Wright (1992: ch. 1; 1996: 923–4; 1999: 225). It should be noted that, in the first instance, Wright uses realism/anti-realism considerations as a motivation for his minimalist framework about truth. Within this framework the concept of truth is characterized by a collection of core principles—"platitudes" in Wright's terminology—that connect truth to other concepts. Minimalism falls short of pluralism. It is possible to adopt a Wright-style principle-based approach to truth and still be a monist. However, Wright thinks that pluralist minimalism overall accommodates the differential appeal of realism and anti-realism better than monist minimalism.

⁶ Edwards (2011b; 2013); Lynch (2009: 57–8); and Wright (2013) all use an analogy between winning and truth. None presents the analogy with winning as a motivation for pluralism. Instead the analogy with winning is meant to introduce—or shed light on—their preferred brand of pluralism. However,

A third motivation for truth pluralism goes via semantic paradox. Roughly speaking, semantic paradoxes are truth-related paradoxes. The most famous semantic paradox is the so-called liar paradox. The paradox is generated by considering the following sentence:

(L) This sentence is not true.

(L) is true if and only if it is not true—hence, the paradox. The liar paradox and other semantic paradoxes show that we must be careful in how we think about truth. Interestingly, it has recently been argued that truth pluralism provides a good way to deal with the liar paradox.⁷ This is not the place to engage in a careful, extensive discussion of this idea. We restrict ourselves to making the following observation: the pluralist solution to the liar paradox gives one reason to take pluralism seriously, assuming that it is a good one. It does so in the sense that one should take seriously views that hold the prospects of providing good solutions to long-standing, difficult problems such as the liar paradox.⁸

20.3 CHALLENGES AND OBJECTIONS

The distinctive thesis of pluralism is that truth is Many. There are several ways of being true. In the previous section we presented three reasons to buy into this thesis. In this section we take a step in the opposite direction by having a look at some reasons why one might find it problematic.

Eklund (2014) suggests that the analogy might carry some weight as a way to motivate pluralism about truth. This is so if pluralism about winning is taken to be uncontroversial and the case of winning and the case of truth are relevantly similar.

⁷ Beall (2013) and Cotnoir (2013a). Note that the kind of pluralism motivated by Beall and Cotnoir is different from the kind of pluralism to be found in the works of Lynch and Wright. Lynch and Wright-style pluralism ties differences in truth to differences in subject-matter. This connection goes missing if we arrive at pluralism through paradox. For in that case, what drives pluralism is not that different subject-matters call for different ways of being true, but rather that a plurality of truth predicates is needed in order to block paradox.

⁸ Whether the pluralist solution to the liar paradox is a good one depends on a careful assessment of the proposals made by Beall and Cotnoir respectively. Also, it is a highly complex matter to give an *overall* assessment of whether one theory of truth fares better than others. The semantic paradoxes make for one relevant factor. However, there are others. Strictly speaking, a given theory's ability to address the liar paradox only gives one a reason to adopt that theory over another theory provided that (i) the former theory provides a *better* solution than the latter, and (ii) the former theory does at least as well overall as the latter theory with respect to other tasks, problems, and issues relevant to assessing the merits of theories of truth.

20.3.1 The instability challenge

A fundamental objection to truth pluralism is that the view is inherently unstable. Pluralists maintain that there are several ways of being true. Truth is Many. However, the following simple line of reasoning—the *instability challenge*—seems to show that truth is really One: assume that there are several ways of being true, as the pluralist would have it. Let them be T_1, \dots, T_n , and characterize the generic way of being true, T_G , by saying that any proposition p is T_G if and only if p is T_1 or \dots or p is T_n . Consider now T_G . Every proposition that is true in some way is true in the T_G -way, regardless of its domain. Furthermore, T_G is co-extensional with the class of propositions that are true in some way. It applies *precisely* to the propositions in that class. We get something even stronger, in fact. T_G is necessarily co-extensional with the class of propositions that are true in some way. At any possible world w , p is T_G precisely if it is true in some way. These observations suggest that T_G is a generic way of being true—a way of being true that applies across the board. However, this shows that truth is really One and not Many, *contra* pluralism. Truth pluralism is thus inherently unstable. If it is assumed to be the correct view, straightforward reasoning reveals a commitment to a generic way of being true—which by its very nature undermines the tenability of the view.⁹

20.3.2 The unity challenge

Suppose that something is *thingy-thingy* if and only if it is divisible by two with no remainder, has two hydrogen atoms and an oxygen atom, is valid, is blue, or loves soccer. There are several ways of being thingy-thingy. Being even, being H_2O , being valid, being blue, and being soccer-loving are all ways of being thingy-thingy. These ways are not ways of being funny, ways of being healthy, or ways of being colored. They are ways of being *thingy-thingy*. What makes it so? Arguably nothing more—or other—than their being properties featuring in the disjunctive characterization of thingy-thingy-ness. To put the point at the level of the objects: the set of thingy-thingies includes even numbers, water molecules, valid arguments, blue objects, and soccer fans. There seems to be nothing genuinely in common between these things. They are unified merely in the sense of all being thingy-thingy; that is, having a disjunctive, gerrymandered property.

The pluralist buys into the thesis that there are several ways of being true, T_1, \dots, T_n . Now, T_1, \dots, T_n are not ways of being funny, ways of being healthy, or ways of being

⁹ The instability challenge is presented in Tappolet (2000). Strictly speaking, in order to capture the idea that ways of being true apply within specific domains, T_G should be characterized in terms of a disjunction of conjunctions saying that p belongs to domain D_1 and is T_1 or \dots or D_n and is T_n .

colored. They are ways of being *true*. What makes it so? We might say that there is nothing more to ways of being true than being T_1 or . . . or being T_n . In that case ways of being true would be like ways of being thingy-thingy. This seems wrong. Ways of being true appear to be genuinely unified. Among other things, they serve as the criterion of correctness of beliefs, whatever their domain. Likewise, ways of being true are preserved by valid inference—again, whatever their domain. Thus, ways of being true and truths are unlike ways of being thingy-thingy and thingy-thingies. They are characterized by genuine unity.

What we might call the “unity challenge” emerges when the pluralist is asked to account for this unity. She might endorse the idea that the various ways of being true are characterized by the kind of core features mentioned above and maintain that this explains their unity or what is in common between them. However, if this line is taken, the pluralist would seem hard pressed not also to grant the existence of a generic way of being true. All ways of being true share certain core features. But sharing these features would appear to make for genuine similarity. In turn, genuine similarity among propositions that correspond, cohere, etc. supports the idea that these propositions share a genuine property. This shared property is not just any old property, though. It would appear to be a generic *truth* property. However, if so, there is a generic way of being true that applies across all domains of truth-apt discourse. Truth is One. This would seem to run counter to the pluralist thesis that truth is Many.¹⁰

20.3.3 Mixed discourse: compounds and inferences

Mixed discourse is discourse that involves more than one domain. To illustrate:

- (M-COMP) Some mountains have snow-covered tops, and breaking and entering is illegal.
- (M-INFER) (I) If some mountains have snow-covered tops, then breaking and entering is illegal.
 (II) Some mountains have snow-covered tops.
 (III) So, breaking and entering is illegal.

¹⁰ What we have here called the “unity challenge” does not feature, as formulated, in Lynch’s work. However, his emphasis on the need to account for the unity of truth can be seen as pointing heavily in its direction (2009: Introduction, ch. 1; 2013: sections 1–2). Both the instability challenge and the unity challenge put pressure on the idea that there is no generic way of being true. However, they do so in different ways. The instability challenge goes through (necessary) co-extension, while the unity challenge zooms in on the idea that, due to a number of shared core features, all ways of being true and all true propositions are genuinely unified.

Mixed discourse has been thought to land the pluralist in trouble on at least two levels: compounds and inference.¹¹ Let us see why.

(M-COMP) could be called a “mixed compound” proposition. It is a mixed conjunction, with one conjunct pertaining to the empirical world and another pertaining to the law. We clearly want to say that (M-COMP) is true. Both of its conjuncts are true, and standard semantics dictates that any conjunction with true conjuncts be true. What is not so clear, though, is how the pluralist is going to account for the truth of the mixed conjunction.¹² This is the so-called problem of mixed compounds, as put in terms of conjunction. What options does the pluralist have? She could say that the conjunction as a whole is true in whatever way the first conjunct is. If correspondence holds sway over discourse concerning snow-covered mountaintops, this means that the entire conjunction—not just the first conjunct—is true in the correspondence way. The pluralist could also say that the conjunction as a whole is true in whatever way the second conjunct is. If coherence with the body of law governs legal discourse, the entire conjunction—not just the second conjunct—is true in the coherence way. Both of these options are unattractive. They violate a basic thought about true conjunctions, *viz.* that the conjuncts contribute equally to the truth of the compound. This thought is violated because, on both proposals, the truth of one conjunct is completely neglected. Saying that (M-COMP) is true in the correspondence way completely neglects the contribution made by the second conjunct’s cohering. Conversely, the suggestion that (M-COMP) is true in the coherence way completely neglects the contribution made by the first conjunct’s corresponding. Now, if (M-COMP) is true neither in the way of the first conjunct nor in the way of the second conjunct, it must be true in some third way. But what way would that be? The pluralist cannot say that (M-COMP) is true in a generic way. This seems to militate against the idea that truth is Many. Three options have been excluded. The task remains for the pluralist to specify in what way mixed conjunctions—and mixed compounds more generally—are true.

(M-INFER) is a mixed inference, an inference whose premises and conclusion cut across the empirical world domain and the legal domain. (M-INFER) is a valid inference. The step from (I) and (II) to (III) is an instance of *modus ponens*—a valid rule of inference. The standard definition of validity is in terms of necessary truth preservation: an argument is valid if and only if, necessarily, if the premises of the

¹¹ So-called mixed atomics have also been thought to cause problems for pluralism. We leave out this kind of mixed discourse for present purposes. Sher presents the problem of mixed atomics in (2005: 321–2). See Lynch (2005b: 340–1); Pedersen and Wright (2012: section 4.5.1); and Wyatt (2013) for pluralist responses.

¹² Tappolet (2000); Williamson (1994). Tappolet does not merely challenge the pluralist to give an account of the truth of mixed compounds, whatever it might be. She further argues that the need to invoke a third way of being true—applicable to the conjunction as a whole—eliminates the need for the two conjunct ways of being true. For this is simply to say that there is just one way of being true, at least as far as conjunctions go.

argument are true, then so is the conclusion. The challenge for the pluralist is how to account for the validity of mixed inferences like (M-INFER).¹³ The talk of truth *preservation* invites the idea that validity involves some property—truth—being passed on or transmitted from the premises to the conclusion. However, within a pluralist framework it seems difficult to sustain this idea. Arguably, (I), (II), and (III) are assessable in terms of distinct ways of being true, and if so, there would seem to be no common way of being true that can be preserved through the inference. The pluralist might try to say that there is a generic way of being true that gets preserved by valid inference. However, as in the case of mixed compounds, taking this path would seem not to be an option—again, it would seem to run counter to the pluralist idea that truth is Many.

20.3.4 The Double-Counting Objection

Many pluralists think that their view provides the best way of accommodating realist and anti-realist sympathies with respect to different domains and also the idea that many different kinds of discourse are truth-apt. Now, suppose that some alternative view was capable of accommodating these phenomena—the realist/anti-realist combination and wide-ranging truth-aptitude—and, also, that the view was more simple than pluralism. In that case, other things being equal, the alternative view would be preferable on grounds of simplicity. Are there any such views? Some deflationists argue that their view is indeed of this kind.

Truth pluralism posits two kinds of difference. There are different domains, and there are different ways of being true. The former wedds the pluralist to plurality at the level of subject-matter, the latter to plurality at the level of how propositions are true. The deflationist contends that buying into differences both at the level of subject-matter and at the level of truth is unnecessary for adequately accommodating wide-ranging truth-aptitude and the differential appeal of realism and anti-realism. According to the deflationist, in both cases, it suffices to acknowledge differences in subject-matter and pair these differences with a single deflationary truth property: one that is exhaustively characterized by the Equivalence Schema or some such schema. When we talk about diverse subject-matters such as mathematics, morals, or the empirical world, there may be differences in the nature of the relevant subject-matters. That is, the things themselves—numbers, wrongness and rightness, and trees and mountains—may be very different kinds of thing. However, these differences at the level of subject-matter are not reflected by corresponding differences in how representations used to talk and think about them are true. The deflationist thus charges the pluralist with what might aptly be called the

¹³ Tappolet (1997); Sainbury (1996).

“double-counting objection”: the pluralist counts two kinds of difference where only one is needed.^{14,15}

20.4 VARIETIES OF TRUTH PLURALISM

At the outset of this chapter truth pluralism was presented as the view that there are several ways of being true. This is only a rough characterization, however. It can be understood in two considerably different ways, giving rise to a distinction between strong and moderate forms of pluralism. This section introduces different versions of these two brands of pluralism.

20.4.1 Core principles and the concept of truth

All pluralists must say how they identify or demarcate putative truth properties. Most pluralists share a methodology on this point—one with an affinity to familiar functionalist theories of property identification. They converge on the following thesis:

(PRIN) The concept of truth (TRUTH) is characterized by a collection of core principles P_1, \dots, P_n that connect TRUTH to other concepts.

Someone who commits to (PRIN) can be regarded as providing a *network analysis* of TRUTH. The endorsement of core principles P_1, \dots, P_n locates TRUTH within a network of

¹⁴ See Blackburn (1998; 2013) and Sainsbury (1996) for this kind of objection. Blackburn (2013: 265) uses the expression “double counting.” What ultimately drives the double-counting objection is the idea that a simpler view on truth is to be chosen over a less simple one, assuming that the simpler view does as well (or better) than the less simple view with respect to all relevant explanatory tasks (2013: 263–4). Traditional monists such as the correspondence theorist might also try to run the double-counting objection against the pluralist. After all, they may well acknowledge differences at the level of subject-matter, but they will deny that they go with differences at the level of truth. However, if confronted with the double-counting objection by a traditional monist, the pluralist would respond by maintaining that the traditional monist has a difficult time accommodating the idea of wide-ranging truth-aptitude. Arguably, the deflationist is in a better position to run the double-counting objection. It is difficult for the pluralist to deny that the deflationist is entitled to claim wide-ranging truth-aptitude because the very principle that the deflationist takes to be characteristic of truth—the Equivalence Scheme or some similar schema—is one that the pluralist will accept for each truth-apt domain. For other simplicity-based considerations against pluralism, see Dodd (2002; 2013); Jackson (1994); Pettit (1996).

¹⁵ The challenges and objections presented in this section by no means cover all challenges or objections that have been launched against the view. Further challenges include the Problem of Generalization (Lynch 2004b: 389; 2009: 57), the Criteria Problem (Wright 2005), the Problem of Epistemic Circularity (Wright 2010), the Problem of Normativity (Lynch 2004b: 389–91; 2009: 57–9), and the “problem of death by a thousand cuts” (Wright 2012).

concepts and takes TRUTH to be characterized by the relations it bears to other concepts in the network.¹⁶

Different pluralists have suggested different principles as candidates for core principles. Lynch (2009; 2013), for example, adopts the following principles:

- (O) For every proposition p , p is true if and only if, were p to be believed, things would be believed to be as they are. (*Objectivity*)
- (NB) For every proposition p , it is prima facie correct to believe p if and only if p is true. (*Norm of Belief*)
- (EI) For every proposition p , other things being equal, if p is true, then believing p is a worthy goal of inquiry.¹⁷ (*End of Inquiry*)

Wright (1992; 1998a; 2001) suggests a different set of principles.

Although advocates of principle-based pluralism endorse distinct sets of core principles, their shared commitment to characterizing TRUTH via network analysis generates further significant common ground: conceptual monism about TRUTH. Principle-based pluralists take their favored core principles to pin down a unique concept, TRUTH. Since most pluralists buy into some version of principle-based pluralism, most pluralists are monists about TRUTH.

20.4.2 Strong and moderate truth pluralism

Let us return to the idea that there are several ways of being true. Strong truth pluralism (STP) and moderate truth pluralism (MTP) differ over how they add detail to this idea:

- (STP) There are several truth properties and, among them, there is no single property that applies across the full range of truth-apt discourse. Truth is constituted by or reduces to different properties within different domains.
- (MTP) There is a single truth property—truth-as-such—that applies across all truth-apt domains of discourse. Propositions belonging to different domains possess this property in virtue of possessing different properties.

Strong truth pluralism and moderate truth pluralism both incorporate the idea that there are several ways of being true. However, they do so in different senses. Strong

¹⁶ This kind of approach is not, of course, specific to the literature on truth. See e.g. Jackson (1998) and Lewis (1979).

¹⁷ The formulation used here is the regimented version suggested by David (2013).

pluralism involves a pluralization of truth itself: truth is constituted by or reduces to correspondence, coherence, or some other property within specific domains. As a bare minimum, the label “truth-as-such” should be applied only to properties that span the full range of truth-apt domains. The strong pluralist gives up on the very idea of truth-as-such—at least at the level of properties. According to strong pluralism, truth is Many—and just that.¹⁸

Strong pluralism is not a widely held view. Early critics of Wright suggested that truth pluralism—as presented in Wright (1992)—carries a commitment to the thesis that “true” is ambiguous. Wright forcefully rejected this suggestion.¹⁹ Many other pluralists follow Wright in rejecting any version of pluralism that is wedded to ambiguity because it is a form of strong pluralism. As such, strong pluralism is thought to run into a whole host of problems—some of which we will detail below.²⁰

Polysemy pluralism

According to *polysemy pluralism*, “true” is polysemous, i.e. it expresses distinct—yet related—meanings.²¹ Sometimes the word picks out a deflationary truth concept, TRUTH_D , and sometimes it picks out an inflationary one, TRUTH_I . TRUTH_D is exhaustively characterized by the Disquotational Schema (“ p ” is true if and only if p). The property of being true in the deflationary way is the property of this concept. TRUTH_I is characterized by the Disquotational Schema together with an objectivity requirement. As a bare minimum, this requirement says that anyone who judges $\sim p$ is at fault if someone else correctly judges p . The property of being true in the inflationary way is the property of the inflationary concept. In sum, according to polysemy pluralism, we have the following combination: one predicate, several concepts, several properties.

Univocal strong pluralism

According to univocal strong pluralism, “true” is relevantly similar to univocal, flexible descriptions. Consider for example the definite description “the winner of the Champions League.” This description is univocal. The description always means something like this: *the team that scores the most goals in the final of the Champions League*. However, the description is flexible. It does not pick out the same team across all contexts. As used in the summer of 2008 the description picked out Manchester United,

¹⁸ It is important to keep separate commitments at the level of concepts and commitments at the level of properties. One might be a conceptual monist without buying into a corresponding thesis at the level of properties. That is, one might think that there is a unique truth concept without thinking that there is a corresponding truth property. Indeed, some strong pluralists fall into this category. What is called “univocal strong pluralism” in this section is an example of this type of view. Moderate pluralists, on the other hand, take the core principles to pin down a unique concept of truth, *and* they think that this commitment to monism filters through to the level of properties. The features of TRUTH —i.e. the concept—mark features of a corresponding property of being true-as-such.

¹⁹ See Sainsbury (1996); Pettit (1996); Wright (1996).

²⁰ Lynch (2004; 2006; 2009).

²¹ Kölbel (2008) and (2013).

while, as used in the summer of 2014, it picked out Real Madrid. Move on to truth. According to *univocal strong pluralism*, there is a set of core principles that connect truth to other concepts. These core principles characterize a unique concept, TRUTH. “True” always expresses this concept and for this reason “true” is univocal, not polysemous. However, while the meaning of “true” is stable across contexts, several properties serve as the referent of “true”. “True” as it pertains to empirical world discourse might refer to correspondence, while “true” as it pertains to legal discourse might pick out coherence with the body of law. Among these various properties, there is no single property that applies across the board. In sum, we have the following combination: one predicate, one concept, several properties. Thus, univocal strong pluralism is a form of pluralism less radical than polysemy pluralism. Polysemy pluralism involves pluralization at two levels—the level of concepts and the level of properties. Univocal strong pluralism only involves the latter kind of pluralization.²²

Let us now turn to moderate pluralism. Moderate pluralists converge on the idea that truth is both One and Many—or, more carefully put, truth is One and its grounds are Many. Truth is One because there is a single property—truth-as-such—that applies across the board. The grounds of truth are Many because, for different domains, truth-as-such is instantiated by propositions in virtue of their instantiating one among a plurality of properties. Moderate pluralists also share the commitment to principle-based pluralism and are monists about the concept of truth. A collection of core principles characterizes a unique concept, TRUTH, and “true” always expresses this concept. We thus have the following combination of commitments: one truth predicate (“true”), one truth concept (TRUTH), one truth property (truth-as-such), and several truth-grounding properties (correspondence, coherence, superassertibility, etc.). These are core commitments of all forms of moderate pluralism on the market. Let us consider moderate pluralism further by investigating the following two aspects of four varieties of the view: (a) the characterization of truth-as-such, and (b) the account of the relationship between the One and the Many, i.e. between truth-as-such and domain-specific properties such as correspondence, coherence, and superassertibility. We consider (a) and (b) in connection with second-order functionalism, manifestation functionalism, determination pluralism, and alethic disjunctivism. For present purposes, we take Objectivity, Norm of Belief, and End of Inquiry to be the core principles that characterize the truth concept. However, nothing in our exposition rests on this.²³

²² Lynch (2006: section 2; 2009: 59–62) uses univocal, non-rigid descriptions to introduce univocal strong pluralism. He does so in the context of a discussion of Wright-style pluralism and attributes univocal strong pluralism to Wright, as represented by Wright (1992; 1996; 1998; 2001). Versions of univocal strong pluralism are developed and defended in detail in Ferrari, Moruzzi, and Pedersen (Ms); Kim and Pedersen (2018); and Pedersen (Ms-c). Pedersen (2006) was intended to be a version of univocal strong pluralism although no explicit mention was made of core principles. Cotnoir (2009; 2013a; 2013b) also advocates strong pluralism although, as in the case of Pedersen (2006), there is no explicit discussion of core principles and the issue what role, if any, they play.

²³ Gila Sher and Bob Barnard and Terry Horgan have developed versions of the correspondence theory that are similar to moderate truth pluralism. Both Sher and Barnard and Horgan think that truth

Second-order functionalism

According to second-order functionalism, truth is a second-order functional property.²⁴ On this view, for a domain D_i and a property F_i , F_i plays the *truth-role* for D_i if and only if F_i satisfies the core principles for D_i -propositions. That is, F_i serves a certain function. Taking Objectivity, Norm of Belief, and End of Inquiry as the core principles this means that a property plays the truth-role for a given domain precisely if it satisfies these three principles with respect to propositions of that domain. Thus, correspondence plays the truth-role for discourse about macroscopic physical objects precisely if, within this domain, things would be as they are believed to be just in case they correspond; beliefs are *prima facie* correct if and only if they correspond; and believing p is a worthy goal of inquiry just in case p corresponds. Similarly for other domain-property pairings. Now, letting p be any proposition and D_i the domain to which p belongs, p is *true-as-such* if and only if p has the property that plays the truth-role for D_i . The property of being true-as-such is the property of having the property that plays the truth-role for the target domain. We see that no matter what domain a proposition belongs to it is truth-as-such—not some other property—that counts as truth. Truth-as-such is a generic truth property. Note that a range of compound-specific properties play the truth-role for logical compounds. For example, for conjunctions the property that plays the truth-role is the property of being a conjunction with conjuncts that all have the property that plays the truth-role for their respective domains.²⁵

There is an intimate relationship between the One and the Many. The relationship is one of realization. Properties such as correspondence, coherence, and superassertibility play the truth-role relative to specific domains, and within these domains they realize truth-as-such. Since different properties realize truth relative to different domains, truth-as-such is multiply realizable. Multiple realizability is not identity. Thus, the relationship between the One and the Many is not intimate in the very strong sense that the latter eliminates the former. Truth-as-such is a property in its own right. Nonetheless, truth-as-such depends asymmetrically on its realizer properties. Truth-as-such is instantiated by a given proposition *because* the proposition instantiates the property that plays the truth-role for the domain to which it belongs.²⁶

is always correspondence. Thus, there is such a thing as true-as-such or generic truth. However, they likewise buy into the idea that different forms of correspondence apply to different kinds of discourse. For further details concerning correspondence pluralism, see Sher (2005; 2013; 2015; 2016); Horgan (2001); Barnard and Horgan (2006; 2013).

²⁴ Second-order functionalism has been developed in detail by Lynch. See Lynch (2000; 2001; 2004a; 2004b; 2005a; 2005b; 2006). Pettit (1996: 886) presents a functionalist proposal in passing. See also Devlin (2003).

²⁵ Lynch (2004b: 397).

²⁶ Wright occasionally uses the expression “variable realization” (1996: 924; 1998a; 2001: 761). It is worth noting that Lynch (2006; 2009) takes Wright (1992; 1994a; 1994b; 1996; 1998a; 2001) to endorse a strong version of pluralism. Wright (2013) expresses sympathy toward a moderate form of pluralism. See Wright (2005; 2010) for criticisms pertaining to Lynch’s principle-based approach. Nulty (2010) and Horton and Poston (2012) object to second-order functionalism on metaphysical and explanatory grounds.

Manifestation functionalism

According to manifestation functionalism, the property of being true-as-such is the property that, as a matter of conceptual necessity, has the features picked out by the core principles—or which plays the truth-role *as such*.²⁷ And a property plays the truth-role as such if and only if, necessarily, it satisfies the three core principles.

How does the single property of truth-as-such relate to the properties that determine or ground it? Following Yablo (1992) and Shoemaker (2001) in the philosophy of mind, the manifestation functionalist takes functional properties to be similar in some respects to determinable properties and characterizes the metaphysical relation in question as manifestation. Manifestation is characterized as follows: a property *F* manifests a property *F** if and only if the essential features of *F** is a subset of *F*'s features.²⁸ Within specific domains these various properties manifest truth-as-such. This is because the features marked by the core principles—i.e. the truisth ones—are included among the features of correspondence, superassertibility, and whatever other properties manifest truth-as-such within specific domains. Note that manifestation is reflexive. Thus, not only do correspondence and other domain-specific properties manifest truth-as-such. All of these properties also manifest themselves. Truth-as-such manifests truth-as-such, correspondence manifests correspondence, and so on.²⁹ Truths that have two truth-manifesting properties—truth-as-such and some other property—are *unplain truths*. Truths that have only one truth-manifesting property—truth-as-such itself—are *plain truths*. Logical compounds fall into the category of plain truths. The only truth-manifesting property that they possess is truth-as-such itself.³⁰

Determination pluralism

Another form of moderate pluralism is so-called determination pluralism.³¹ As with manifestation functionalism, truth-as-such is the property that necessarily satisfies the core principles for every (truth-apt) proposition.³² But on this account, truth is said to be only metaphysically determined by its grounding properties. The conditions for such

²⁷ Lynch (2009: 74).

²⁸ Lynch (2009: 74).

²⁹ The characterization of manifestation given here is simpler than Lynch's own (which involves talk of the a priori and features being *conceptually* essential). For present purposes this does not matter. Lynch is keen to present manifestation as a new metaphysical relation, one different from the relation that obtains between a genus and its species (2009: 67) and between a determinable and its determinants (2009: 75). See Tappolet (2010) and Pedersen (Ms-a) for critical discussion.

³⁰ Lynch (2009: 90–91; 2013: section 6). In his (2009) Lynch takes the atomic/compound distinction and the unplain/plain truth distinction to coincide. Shapiro (2011) contests this idea by arguing that there are plain, atomic truths (e.g. truth-attributions). Lynch (2013: section 6) responds to Shapiro.

³¹ Edwards (2011b; 2013).

³² Edwards (2011b: 40).

determination are given by certain (necessarily true) conditionals. To see the strategy here, compare how we characterize winning for different sorts of games:

(WIN) In G : if a player/team has property F , then that player/team has won.

Here G is a game and F is the property that triggers a win in G . Whichever player/team has property F wins. F determines the outcome of the game, and in this sense F is a winning-determining property. Sample winning-determination conditionals include:

(WIN-S) In soccer: if a team has the property of having scored more goals, then that team has won.

(WIN-C) In chess: if a player has the property of having checkmated the opponent's king, then that player has won.

In soccer the winning-determining property is the property of having scored more goals. If a team has the property of having scored more goals, this determines the outcome of the game: the team wins. In chess the winning-determining property is the property of having checkmated the opponent's king. If a player has the property of having done so, this determines the outcome of the game: that player wins.

Determination pluralists think that truth works in the same way. Truth within specific domains can be characterized by a conditional of the following form:

(TD) In D : if p is F , then p is true-as-such.

Here D is a domain, p a proposition, and F the property that is "alethically potent" within D . Sample truth-determination conditionals include:

(TD_{MPO}) In the domain of macroscopic physical objects: if p corresponds to reality, then p is true-as-such.

(TD_M) In the domain of morals: if p is superassertible, then p is true-as-such.

The determination pluralist takes the relationship between the One and the Many to be one of determination. In discourse concerning macroscopic physical objects correspondence determines truth: if a proposition corresponds to reality, it is true-as-such. Superassertibility stands in the same relation to moral discourse.³³

³³ Wright (2013) expresses sympathy toward moderate pluralism, especially determination pluralism.

Alethic disjunctivism

A final form of moderate pluralism is alethic disjunctivism.³⁴ Suppose that there are n truth-apt domains, D_1, \dots, D_n . Let F_1, \dots, F_n be the properties that satisfy the core principles for propositions belonging to these domains. Characterize truth-as-such disjunctively, as follows:

(T_v) Any proposition p is true-as-such if and only if p is F_i and belongs to D_i or \dots or p is F_n and belongs to D_n .

Each of the domain-specific properties F_1, \dots, F_n necessarily satisfies the core principles relative to their respective domains. On this assumption it can be shown that the disjunctive property defined by (T_v) also satisfies the core principles. Indeed, stronger yet, it can be established that it does so necessarily.³⁵

Consider a domain D_i , a proposition p pertaining to D_i , and let F_i be the property that is alethically potent within D_i . In that case, p is true-as-such if and only if p is F_i . The relationship between truth-as-such and domain-specific properties such as F_i is one of grounding. Thus, the bi-conditional relationship between truth-as-such and F_i should be read with priority from right to left: p 's being F_i grounds p 's being true-as-such. p is true-as-such *because* p is F_i .³⁶

General remarks

To wrap up this section let us briefly record two related points concerning the One–Many relations favored by the four kinds of moderate pluralist considered above. First, note that there is an intimate relationship between the One and the Many. Instantiations of the domain-specific properties trigger instantiations of truth-as-such. Second, related to the first point, note that the relations of realization, determination, and grounding must be relations of asymmetric dependence to serve the purposes of moderate pluralism. (A relation R is strongly asymmetric if and only if, for all x, y , if Rxy , then $\sim Ryx$.) Let T_G be truth-as-such, and let F_i be the domain-specific property relevant to p 's domain. In that case asymmetry tells us that

If $F_i(p)$ realizes $T_G(p)$, then it is not the case that $T_G(p)$ realizes $F_i(p)$.

If $F_i(p)$ determines $T_G(p)$, then it is not the case that $T_G(p)$ determines $F_i(p)$.

If $F_i(p)$ grounds $T_G(p)$, then it is not the case that $T_G(p)$ grounds $F_i(p)$.

³⁴ Edwards (2012); Pedersen (2006; 2010; 2012a); Pedersen and Edwards (2011); Pedersen and Wright (2013b).

³⁵ See Pedersen (2012a) for details. See also Pedersen and Edwards (2011) and Pedersen and Wright (2013b). It is interesting to note that, by their own lights, manifestation functionalists and determination pluralists should arguably also count the disjunctively characterized property as a candidate for being truth-as-such.

³⁶ Pedersen and Wright (2013b).

If attention is restricted to instances involving distinct properties, manifestation is also asymmetric, and so, in particular,

If $F_i(p)$ manifests $T_G(p)$, then it is not the case that $T_G(p)$ manifests $F_i(p)$.³⁷

The Many ground the One, not the other way around. Instantiations of truth-as-such asymmetrically depend on instantiations of properties such as correspondence and superassertibility.

20.5 MODERATE PLURALISM: RESPONSES TO CHALLENGES AND OBJECTIONS

We now turn to the issue how pluralists might address the challenges presented in section 20.3. As noted earlier, between moderate pluralism and strong pluralism the former is the more widely held view—and this by a wide margin. This is largely due to the fact that, while moderate pluralists have offered responses to most challenges and objections, not much of an attempt has been made to defend strong pluralism. In this section we discuss moderate pluralist responses to the challenges and objection presented in section 20.3. Section 20.6 presents responses on behalf of strong pluralists.

20.5.1 Response to the instability challenge

The instability challenge is meant to show that basic logical reasoning takes one from a commitment to several ways of being true to a generic way of being true. Moderate pluralists can respond to the instability challenge by noting that its force rests on the assumption that pluralists cannot embrace a generic way of being true. However, while this is true of strong pluralists, moderate pluralists are not constrained in the same way. Indeed, their view is characterized precisely by the accommodation of a generic truth property. Thus, as seen in the previous section, second-order functionalists, manifestation functionalists, determination pluralists, and alethic disjunctivists all endorse a truth property that applies across all truth-apt discourse—generic truth, that is. Making

³⁷ If we do not restrict the claim concerning asymmetry to distinct properties, inconsistency results. This is because manifestation is reflexive, and so, any property manifests itself. Suppose that $T_G(p)$. By reflexivity, $T_G(p)$ manifests $T_G(p)$. Now, if manifestation was generally asymmetric, we would have the following instance: if $T_G(p)$ manifests $T_G(p)$, then it is not the case that $T_G(p)$ manifests $T_G(p)$. But given the instance of reflexivity just cited, we would then get that it is not the case that $T_G(p)$ manifests $T_G(p)$ and be stuck with a contradiction. This is why the claim concerning the asymmetry of manifestation is restricted to instances of the relation featuring distinct properties. For critical discussion of manifestation, see Pedersen (Ms-a).

this observation about the commitments of moderate pluralism would appear to be sufficient to defuse the instability challenge.

20.5.2 Response to the unity challenge

The unity challenge puts pressure on the pluralist by pointing out that a number of core features seem to unify the pluralist's ways of being true. Since these features seem to make for genuine similarity, it would seem difficult to resist the conclusion that there is a genuine property shared by all and only the truths—a generic truth property, as it were. This is unfortunate because it seems to run counter to the thesis that truth is Many. In response to this challenge—and as in the case of the instability challenge—moderate pluralists can point out that they explicitly commit to the supposedly problematic generic truth property. Furthermore, moderate pluralists embrace the idea that a range of core features are characteristic of the various domain-specific properties. After all, their view is principle-based. They think that certain core principles characterize the concept of truth, and that these principles mark core features possessed by the various domain-specific properties. Rather than being an objection to their view, it would thus seem that the unity challenge is more of a description of commitments that the moderate pluralists have already.³⁸

20.5.3 Response to challenges concerning mixed discourse

Two problems concerning mixed discourse were presented in section 20.3: the problem of mixed compounds and the problem of mixed inferences. Let us start by pointing to a line of response that does not work. Doing so serves as a good introduction to lines of response that hold a greater promise of being adequate.

Moderate pluralists are committed to saying that any true proposition, whether pure or mixed, is true-as-such. For this reason they are also committed to saying that the premises and the conclusion of any valid inference—whether pure or mixed—are true-as-such. They are thus in a position to say that mixed inferences preserve truth-as-such in the sense that the premises and the conclusion of any such inference share this property. Merely making these observations, however, does not work as a line of response to the problems concerning mixed discourse. It does not work because these problems are not only problems urging pluralists to show *that* they can point to a truth property possessed by mixed compounds and that gets preserved in mixed inferences (in the sense that it is shared by both premises and conclusion). The *real* problem of mixed compounds is to account for *how* mixed compounds get to be true.

³⁸ Something like the unity challenge can be regarded as fundamentally shaping and informing the work of both Lynch and Wright. See Lynch (2009; 2013) and Wright (1992; 1998; 2001; 2013).

Merely being in a position to say *that* they are so is not enough. If the problem of mixed inferences is taken to concern truth preservation, the *real* problem of mixed inferences is to account for *how* they preserve some truth property. Again, merely being able to say *that* some truth property gets preserved does not suffice on its own. Within the framework of moderate pluralism, taking on board the points just made, we see that the *real* problem of compounds amounts to accounting for *how* or *why* mixed compounds are true-as-such. It does not suffice merely to observe *that* they are so. The *real* problem of mixed inferences amounts to accounting for *how* truth-as-such is preserved from premises to conclusion.³⁹

20.5.3.1 *Mixed compounds*

We now turn to the problem of mixed compounds. We discuss the problem in relation to second-order functionalism first and then move on to manifestation functionalism, determination pluralism, and alethic disjunctivism—in that order.

Mixed compounds and second-order functionalism

The idea behind second-order functionalism is that propositions are true-as-such because they have the property that plays the truth-role for their domain. We noted earlier that there is a range of compound-specific properties that play the truth-role for logical compounds such as conjunctions and disjunctions. This means that mixed compounds do not pose any special problem. For any type of logical compound the truth of mixed compounds of that particular compound type is accounted for in precisely the same way as the truth of pure compounds of that type. Consider conjunction. The property relevant to this type of compound is the property of being a conjunction with conjuncts that all have the property that plays the truth-role for their domain. A pure mathematical conjunction such as $\langle 2 + 2 = 4 \text{ and } 98 - 1 = 97 \rangle$ is true-as-such because it is a conjunction with conjuncts that have the property that plays the truth-role within their domain. For a mixed conjunction such as $\langle \text{Some mountains have snow-covered tops, and breaking and entering is illegal} \rangle$ the story is the same. There is no special problem of mixed compounds.

Mixed compounds and manifestation functionalism

The manifestation functionalist draws a distinction between two kinds of truths—plain and unplain ones. Recall that all properties self-manifest. In particular, truth-as-such

³⁹ The points just made are often missed by moderate pluralists in their discussion of challenges raised by mixed discourse. See Pedersen (Ms-c); Pedersen (Ms-d); Ferrari, Moruzzi, and Pedersen (Ms); Kim and Pedersen (Ms) for further details and discussion. The conditional formulation used in connection with the problem of mixed inferences is meant to indicate that one does not necessarily have to think of validity in terms of preservation of a single property. For a sketch of a solution (due to Pedersen) to the problem of mixed inferences that rejects this assumption, see section 20.6.6.

manifests itself. This means that any truth will have at least one truth-manifesting property, truth-as-such. Plain truths are those truths for which truth-as-such is the only truth-manifesting property. Unplain truths are those truths that have two truth-manifesting properties, truth-as-such and some other property. \langle There are mountains \rangle is unplainly true. Supposing that correspondence is the truth-manifesting property for discourse about macroscopic physical objects, we see that \langle There are mountains \rangle has two truth-manifesting properties: corresponding and truth-as-such. Any compound, on the other hand, is plainly true. Thus, both $\langle 2 + 2 = 4$ and $98 - 1 = 97 \rangle$ and \langle Some mountains have snow-covered tops, and breaking and entering is illegal \rangle are plainly true. The only truth-manifesting property that they possess is truth-as-such. As with second-order functionalism, mixed compounds raise no special problem. They receive the exact same treatment as pure compounds.

It is worth noting that the manifestation functionalist's account of compound truth points to a difference in how atomic truth and compound truth are grounded. Atomic truths are strongly grounded. They are true-as-such because they possess some truth-manifesting property distinct from truth-as-such. For example, \langle There are mountains \rangle is true-as-such because it corresponds. However, since compounds possess no truth-manifesting property distinct from truth-as-such, they cannot be strongly grounded. Instead they are weakly grounded: their truth supervenes on the unplain truth of their atomic parts.⁴⁰

Mixed compounds, determination pluralism, and alethic disjunctivism

As presented in section 20.4, at least superficially, neither determination pluralism nor alethic disjunctivism is equipped to deal with mixed compounds. This is because they both appear to be incomplete views—and for the same kind of reason. Determination pluralism and alethic disjunctivism are alike in that they connect truth-as-such to domain-specific properties on the basis of conditionals of the form:

(D_i) If p belongs to domain D_i , then if p is F_i , then p is true-as-such.

(D_j) If p belongs to domain D_j , then if p is F_j , then p is true-as-such.

Here F_i and F_j are the truth-relevant properties for domains D_i and D_j respectively.⁴¹ Note that (D_i) and (D_j) are domain-specific conditionals. As such, they seem to remain silent on how the truth-as-such of compounds is connected to domain-specific properties. However, it turns out that determination pluralism and alethic disjunctivism pack in enough information to tell a story about compounds. A bit of footwork must be done to unpack this information, though. Let us sketch how this can be done.

⁴⁰ Lynch (2009: 90–1; 2013: section 6) draws the distinction between strong and weak grounding. For an extensive criticism of the metaphysics of manifestation functionalism, see Pedersen (Ms-a).

⁴¹ Alethic disjunctivism involves the following disjunctive characterization of truth-as-such: any proposition p is true-as-such if and only if p is F_i and belongs to domain D_i , or ... or p is F_n and belongs to domain D_n . It is easy to show that this characterization entails the kind of conditionals in question.

Consider the following principles:

- (CON) If p belongs to D_i and q belongs to D_j , then if p is F_i and q is F_j , then $p \wedge q$ is true-as-such.
- (DIS) If p belongs to D_i and q belongs to D_j , then if p is F_i or q is F_j , then $p \vee q$ is true-as-such.
- (\wedge I) If p is true-as-such and q is true-as-such, then $p \wedge q$ is true-as-such.
- (\vee I) If p is true-as-such or q is true-as-such, then $p \vee q$ is true-as-such.

For (CON) and (DIS), when $D_i = D_j$ (and thus, $F_i = F_j$), we are dealing with a pure compound. When D_i and D_j are distinct, we are dealing with a mixed compound. (CON) and (DIS) are conditionals that do for, respectively, conjunctions and disjunctions what (D_i) and (D_j) do for atomics—namely, connect the truth-as-such of the target proposition to truth-relevant properties. There is one difference, though. Atomic propositions connect to truth-as-such by having the truth-relevant property of their domain. Compounds, on the other hand, connect to truth-as-such by their constituents' having the requisite semantic status. (\wedge I) and (\vee I) are standard semantic clauses for conjunction and disjunction, rewritten in terms of truth-as-such so as to fit our present framework.

Looking at (CON) and recalling that the determination pluralist favors the determination relation, what the determination pluralist would want to say is that each conjunct's possession of its truth-relevant property is individually necessary and jointly sufficient to determine the truth-as-such of the conjunction. For (DIS) what the determination pluralist would want to say is that one disjunct's possession of its truth-relevant property is sufficient to determine the truth-as-such of the disjunction. Recalling that the alethic disjunctivist favors the grounding relation, the alethic disjunctivist would want to give the same kind of account—although with grounding taking the place of determination. To illustrate how this works let us consider the mixed conjunction ⟨Some mountains have snow-covered tops, and breaking and entering is illegal⟩. Let us suppose that the truth-relevant properties of the first and second conjunct are correspondence and coherence respectively. In that case the determination pluralist would want to say that the first conjunct's corresponding and the second conjunct's cohering is individually necessary and jointly sufficient to determine the truth-as-such of ⟨Some mountains have snow-covered tops, and breaking and entering is illegal⟩. The alethic disjunctivist would want to repeat this—although, again, with grounding taking the place of determination.

For present purposes it is interesting to note that (CON) follows from (D_i), (D_j), and (\wedge I), and that (DIS) follows from (D_i), (D_j), and (\vee I). This makes for an interesting observation because it shows that both determination pluralism and alethic disjunctivism have a story to tell about compounds, whether pure or mixed. Although this story is not explicit in the initial presentation of either view, it can be extracted from conditionals that are included in the statement of the views together with standard semantic

clauses and the favored relation of metaphysical dependence (i.e. determination or grounding).⁴²

20.5.3.2 *Mixed inferences*

In their discussion of mixed inferences moderate pluralists note that both the premises and the conclusion are true-as-such. They think that this suffices to address the problem. However, as noted earlier, this kind of response does not engage with the *real* problem of mixed inferences—the demand for a story that makes it clear why the premises’ being true-as-such makes it the case that the conclusion is true-as-such. Here we offer a partial response to the real problem of mixed inferences. It is partial because, due to constraints of space, we consider only one example of a mixed inference. However, doing so suffices to convey what kind of response can adequately address the real problem of mixed inferences.⁴³

Consider the following mixed inference:

- (I) Some mountains have snow-covered tops.
- (II) It is not the case that some mountains have snow-covered tops, or breaking and entering is illegal.
- (III) Breaking and entering is illegal.

The mixed inference from (I) and (II) to (III) is valid. To account for this suppose that (I) and (II) are both true-as-such. We now have to show that (III) is true-as-such and explain why this is so by reference to (I) and (II). In accounting for the truth-as-such of (I) the moderate pluralist would want to point us to a domain-specific property—correspondence, say. (I) is true-as-such because it corresponds. Turning to (II), by assumption, we again have that it is true-as-such. Now, given the truth-as-such of (I) we know that the first disjunct of (II) is false. Given the semantics of disjunction this means that the second disjunct of (II) has to be true-as-such. Again, the moderate pluralist would want to point us to a domain-specific property—coherence with the law, say. However, this tells us that (III) is true-as-such because it coheres with the law. In sum, what we have is a story that makes it clear *why* (III) is true-as-such if (I) and (II) are. The story achieves this by linking the truth-as-such of (I), (II), and (III) to their truth-relevant properties.

Let us make two remarks before proceeding to the next challenge. First, again, the response just presented is at most partial. Only one kind of mixed inference has been dealt with. Second, although the response is only partial, the general idea should be clear: in order to give an adequate response to the real problem of mixed inferences, moderate pluralists need to show *how* the premises guarantee the truth-as-such of the conclusion.

⁴² The ideas presented in this section are based on Pedersen (2012b: section 5.4) and (Ms-b).

⁴³ This is not the typical response given by moderate pluralists, although, as indicated earlier, responses given by moderate pluralists *ought* to take this form. For further discussion, see Pedersen (Ms-d); Ferrari, Moruzzi, and Pedersen (Ms).

20.5.4 Response to the double-counting objection

The pluralist thinks that there are differences at two levels: at the level of subject-matter and also at the level of truth. The first level is metaphysical, the second semantic. The double-counting objection charges the pluralist with invoking two differences where only one is needed. In order to accommodate the idea of wide-ranging truth-aptness and the appeal of realism and anti-realism with respect to different domains, one only needs to buy into differences in subject-matter. Below we present a pluralist response to this line of reasoning. The response aims to show that it is not feasible to take on board differences in subject-matter without also taking on board differences in truth.

Suppose, with the objector, that there are differences in subject-matter—that for instance macroscopic physical objects are metaphysically different from laws. Consider now the proposition that there are mountains and the proposition that speeding is illegal. Both propositions are true. However, the truth of the former proposition would appear to be a very different circumstance from the truth of the latter. \langle There are mountains \rangle relates to the world in a way significantly different from the way in which \langle Speeding is illegal \rangle does. Mountains are mind-independent entities while laws are social—and so, mind-dependent—constructs. In accounting for the circumstance that \langle There are mountains \rangle is true it seems right to say that there is a fit between the proposition and reality, *and* that this fit is in no way due to us shaping, or somehow contributing to, what the relevant tract of reality is like. Matters change when we turn to \langle Speeding is illegal \rangle . In accounting for the circumstance that this proposition is true, it again seems right to say that there is a fit between proposition and reality. However, this fit is precisely due to us shaping, or somehow contributing to, what the relevant tract of reality is like. We have passed laws that classify speeding as being illegal. \langle Speeding is illegal \rangle thus fits with legal reality because we have made it so. The truth of respectively \langle There are mountains \rangle and \langle Speeding is illegal \rangle is accounted for in significantly different ways, as just seen. But this is just to say that, in addition to there being a metaphysical difference between the two propositions, there is a semantic one. The pluralist does not double count. There are two differences, not just one.⁴⁴

20.6 STRONG PLURALISM: RESPONSES TO CHALLENGES AND OBJECTIONS

Strong pluralism is widely regarded as a non-starter, even within the pluralist camp. The aim of this section is to entertain the possibility that the outright dismissal of strong

⁴⁴ The line of response presented here is a version of the response given by Crispin Wright in an exchange with Blackburn. See Wright (2003a: section IV).

pluralism is premature.⁴⁵ To support this idea the challenges and objections from section 20.3 are briefly considered and responses are sketched on behalf of the strong pluralist. Note that the response given to the double-counting objection in section 20.5 is equally available to the moderate and the strong pluralist, and so this particular objection needs no separate treatment here.

20.6.1 Response to the instability challenge and the unity challenge

Recall that we are assuming that the strong pluralist is a principle-based pluralist. In response to the unity challenge the strong pluralist can thus say that there is unity at the conceptual level: TRUTH—the concept—is unified. It is unified by way of being characterized by a collection of core principles. In turn the truth-relevant properties—the properties that constitute or reduce truth within specific domains—are unified in that they all have the features marked by the core principles. The question remains, though, whether truth is unified at the metaphysical level: whether, that is, there is a generic way of being true or a generic truth property.

At this point the strong pluralist may turn to the idea of explanation and pursue one of two lines of response—one concessive, the other not. The concessive line of response is this: there is a generic way of being true, but it is not explanatorily fundamental or important. To elaborate: suppose that generic truth (T_G) exists. If so, it bears an intimate connection to the various truth-relevant properties (T_1, \dots, T_n) endorsed by the strong pluralist. It is instantiated because some truth-relevant property is: $T_G(p)$ because $T_i(p)$, for some $1 \leq i \leq n$. In light of this, although T_G exists, it is not explanatorily fundamental or important. The truth-grounding properties are more important and fundamental. They ultimately do all the explanatory work. The Many ground the One.⁴⁶

Now consider the non-concessive line of response. Like the concessive response this line of response notes that instantiations of generic truth, T_G , are always grounded in instantiations of the pluralist's favored properties, T_1, \dots, T_n . However, instead of just making a relative claim about explanatory fundamentality or importance, the non-concessive response continues to say that generic truth is *explanatorily dispensable*. There are no explanations for which it is needed. The plausibility of this part of the non-concessive response will depend on whether the strong pluralist can shoulder all explanatory burdens without invoking a generic way of being true (e.g. whether mixed compounds and inferences can be dealt with satisfactorily—more on both of these below). It goes beyond the scope of our brief discussion here to assess whether she can. But suppose

⁴⁵ Some of the most vocal critics of strong pluralism are moderate pluralists. See e.g. Lynch (2004; 2006; 2009) and Wright (2013). Strong pluralism is seriously entertained by one author here (Pedersen) and not so seriously by the other. The ideas presented in this section are drawn from Pedersen (Ms-c; Ms-d); Ferrari, Moruzzi, and Pedersen (Ms); Kim and Pedersen (2018).

⁴⁶ See Pedersen (2010) and Kim and Pedersen (Ms) for this line of response.

that she can and consider the following criterion of ontological commitment: accept the existence of something only if it is explanatorily indispensable. This is a widely endorsed criterion of ontological commitment. It ties ontological commitment to ontological parsimony and explanation. Interestingly, if we assume with the “non-concessivist” that generic truth is explanatorily dispensable, the criterion tells us to reject the existence of generic truth.⁴⁷

20.6.2 Response to the problem of mixed inferences and compounds

20.6.2.1 *Mixed compounds*

The strong pluralist can address the problem of mixed compounds in a way similar to the way in which the second-order functionalist does. For each type of compound, there is a property that is truth-relevant specifically for compounds of that type. In the case of conjunction the property is that of being a conjunction with conjuncts that all have whatever property is truth-relevant for them. According to the second-order functionalist, as seen above, this property is truth-relevant for conjunctions by being truth-grounding, i.e. by being that in which virtue of which conjunctions are true-as-such. However, given her commitments, the strong pluralist buys into a stronger claim: the conjunction-specific property *constitutes* truth for conjunctions.⁴⁸

20.6.2.2 *Mixed inferences*

Validity amounts to a kind of *semantic guarantee*: the positive semantic status of the premises guarantees that the conclusion enjoys a positive semantic status, too.⁴⁹ This is a way to think about validity that does not tie it to the idea of preserving a single property. The semantic-guarantee conception of validity can thus be straightforwardly applied within the framework of strong pluralism. Applying this conception, the problem of mixed inferences becomes the problem of showing how, in a valid mixed inference, the premises’ possession of their respective truth-constituting properties guarantees that the conclusion possesses its truth-constituting property.

As before, consider the following mixed inference:

- (I) Some mountains have snow-covered tops.
- (II) It is not the case that some mountains have snow-covered tops, or breaking and entering is illegal.
- (III) Breaking and entering is illegal.

⁴⁷ For further details, see Kim and Pedersen (Ms).

⁴⁸ For further details see Kim and Pedersen (2018); Pedersen (Ms-d); Ferrari, Moruzzi, and Pedersen (Ms). Other sources that discuss mixed compounds include Edwards (2008; 2009); Cotnoir (2009; 2013b); Cook (2011).

⁴⁹ See Pedersen (Ms-d).

Suppose that (I) and (II) have their truth-constituting properties. This means that (I) corresponds. In turn this means that \langle It is not the case that some mountains have snow-covered tops \rangle does not have its truth-constituting property. However, in light of this, and since (II) has the property of being a disjunction with at least one disjunct that has the truth-constituting property of its domain, \langle Breaking and entering is illegal \rangle must have its truth-constituting property. But this means that (III) coheres with the law. In sum, we see that (I) and (II)'s possession of their truth-constituting properties guarantees that (III) possesses its truth-constituting property.⁵⁰

20.7 PLURALISMS?

So far this chapter has focused exclusively on pluralism about truth. However, does this type of pluralism connect in significant ways with other areas? In particular, does it imply or is it implied by other forms of pluralism? If so, truth pluralism would not be an isolated phenomenon, but rather a member of a cluster of interrelated pluralisms. Traditionally the concept of truth has been treated as standing in intimate relationships with a whole range of other concepts—including, but not limited to, *reference*, *satisfaction*, *meaning*, *validity*, *justification*, *content*, *representation*, *fact*, and *being*. In light of this perhaps we should expect truth pluralism to be a part of a cluster of pluralisms? In this section we briefly introduce several pluralist trends in the literature, and we highlight ways in which different forms of pluralism might be connected.

Let us start by observing that several among the concepts mentioned just above are semantic concepts, partly characterized in terms of one another. For this reason it might be natural to think that truth pluralism features in a cluster of semantic pluralisms. Since the concepts in question are partly defined in terms of one another, if pluralism is held with respect to one of them, one should expect pluralism to be characteristic of the others as well. Not much work has been done to spell out and investigate this kind of cluster of semantic pluralisms. However, it would seem like a natural next point for truth pluralism to move toward.⁵¹

There are several noticeable pluralist trends in the literature. Ontological pluralism has emerged as a topic of interest in metaphysics. Ontological pluralism is the view that there are several ways of being—often understood, along Quinean lines, as saying that there are several fundamental existential quantifiers. A number of people have recently

⁵⁰ In giving the proposed solution to the problem of mixed inferences we help ourselves to the compound-specific truth-constituting properties introduced (in section 20.6.2.1) to deal with mixed compounds. For other strongly pluralist solutions to the problem of mixed inferences, see Beall (2000); Pedersen (2006); Cotnoir (2013b). Lynch (2009: 63) and Cotnoir (2013b) criticize the proposal made in Pedersen (2006). Tappolet (2000) offers a rejoinder to Beall (2000). Wright (2013: section 4) offers a sympathetic elaboration and discussion of Beall's proposal.

⁵¹ In Lynch (2009: ch. 7), a whole chapter is devoted to an investigation of semantic functionalism—a specific version of the kind of cluster of interrelated semantic pluralisms mentioned here. Kölbel (2013: 293–5) discusses whether truth pluralism calls for pluralism about propositions.

developed and defended this view against objections.⁵² Another noticeable pluralist trend pertains to logic. Different versions of logical pluralism have emerged. The type of logical pluralism that has attracted the most attention recently is perhaps the one found in the works of Beall and Restall. *Validity* is necessary truth preservation across cases: an argument is valid_X if and only if, in every case_X, if the premises are true, then so is the conclusion. According to Beall and Restall, there are several equally legitimate notions of validity because case_X can be precisified in several equally legitimate ways. Cases can be construed as worlds (complete and consistent), constructions (potentially incomplete), and situations (potentially inconsistent). These three construals of cases yield classical, intuitionistic, and relevance logic respectively. The literature on Beall–Restall-style logical pluralism and other forms of logical pluralism is steadily growing.⁵³

Let us finish by pointing to ways in which truth pluralism might connect to other kinds of pluralism. Some truth pluralists have argued that pluralism about truth implies a form of pluralism about logic. They reason roughly as follows: validity is defined in terms of truth. Accordingly, if the truth-relevant property within domain D_i goes with a certain logic and the truth-relevant property of some other domain D_j goes with a different logic, arguments pertaining to D_i and arguments pertaining to D_j have distinct logics. Now, for at least one domain, correspondence is the truth-relevant property, and it goes with classical logic. However, for at least one other domain, superassertibility—or some similar anti-realist property—is the truth-relevant property, and it goes with intuitionistic logic. Thus, for different domains the validity of arguments is assessable in terms of different logics. According to this argument pluralism about truth and pluralism about logic are intimately connected. The former supports the latter.⁵⁴ Note that the relevant form of logical pluralism is different from that of Beall and Restall. There is a plurality of domains with different ways of being true and different logics. This is a *domain-based* form of logical pluralism. For Beall and Restall there is a plurality of logics because there is a plurality of equally legitimate notions of case to spell out validity: a *case-based* form of logical pluralism.

One might also think that truth pluralism connects with ontological pluralism. Some truth pluralists have argued that the two kinds of pluralism connect in that a form of ontological pluralism implies truth pluralism. Ontological pluralism is the thesis that

⁵² McDaniel (2009; 2010a; 2010b, 2017); Turner (2010; 2012). Spencer (2012) provides an overview.

⁵³ For details concerning case-based pluralism, see Beall and Restall (2006). They refer to the characterization of validity in terms of necessary truth preservation over cases as “Generalized Tarski’s Thesis”. Field (2009) discusses the issue of whether any form of logical pluralism is both interesting and true. He dismisses a considerable number of forms of logical pluralism as being either false or not particularly interesting. Beall–Restall-style logical pluralism is deemed potentially true but not terribly interesting. Other samples of recent work on logical pluralism include several contributions in Cohnitz et al. (2014); Hjortland (2013); Keefe (2014); Read (2006); Russell (2008); Shapiro (2014). Cook (2010) provides an overview.

⁵⁴ This kind of argument can be found in Lynch (2008; 2009: ch. 5) and Pedersen (2014). Pedersen (Ms-e) criticizes Lynch’s path from moderate truth pluralism to logical pluralism, arguing that moderate pluralists should be logical monists.

there are several ways of being. One way to understand this thesis is against the background of the distinction between mind-dependent and mind-independent existence. Thus understood ontological pluralism is the thesis that some objects exist in a mind-dependent way and other objects in a mind-independent way. The mind-dependence/mind-independence distinction is a historically very prominent one—witness the century-long debate concerning the fundamental nature of reality between those who took reality to be mind-dependent and those who rejected this idea.

There seems to be an argument from ontological pluralism—understood in the above way—to a form of truth pluralism. The argument runs as follows: just as the mind-dependence/mind-independence distinction can be applied at the level of objects, it can be applied at the level of how objects instantiate properties. Doing so leads us to pluralism about property instantiation. There are several ways for objects to instantiate properties: they can instantiate them in a mind-dependent way, and they can instantiate them in a mind-independent way. But now we are close to the thesis that there are different ways of being true. We get there by making a few observations. First, note that we have atomic truths of the form $\langle a \text{ is } F \rangle$ when some object a instantiates a property F . Second, given the first observation, the mind-dependence/mind-independence distinction can be straightforwardly extended to atomic truths in the following manner: an atomic proposition $\langle a \text{ is } F \rangle$ is mind-dependently true if and only if a instantiates F in a mind-dependent way, and $\langle a \text{ is } F \rangle$ is mind-independently true if and only if a instantiates F in a mind-independent way. Lastly, we just need to observe that we have already taken on board the idea that some objects instantiate properties in a mind-dependent way while other objects instantiate properties in a mind-independent way. In sum, the path from ontological pluralism to truth pluralism is as follows: we extend the pluralization of ways of being to ways of being instantiated, and from ways of being instantiated we immediately get ways of being (atomically) true.⁵⁵

In this section we have suggested that truth pluralism may not be an isolated phenomenon—that, indeed, it may be one among several species of pluralism that form an interconnected cluster of views. We have said something about what some of these other forms of pluralism might be and indicated how they might connect to truth pluralism. Much work still remains to be done in terms of investigating these various brands of

⁵⁵ This kind of argument can be found in Pedersen (2014). Pedersen (2014) also goes from truth pluralism to logical pluralism, suggesting that truth pluralism as well as logical pluralism flow from ontological pluralism. For another argument going from ontological pluralism to truth pluralism, see Cotnoir and Edwards (2015). Note that Cotnoir and Edwards also give an argument in the other direction, i.e. from truth pluralism to ontological pluralism. Furthermore, note that the kind of truth pluralism reached through pluralism about property-instantiation falls short of the kind of truth pluralism discussed in sections 20.1–20.6. In earlier sections we connected truth pluralism to properties such as correspondence and superassertibility. The argument from ontological pluralism to truth pluralism given here involved mind-dependent and mind-independent ways of being true. These ways of being true cannot be presumed to be identical to any of the truth-relevant properties mentioned earlier without further argument. See also Yu (2017).

pluralism, adding further kinds to the list (e.g. pluralism about justification), and continuing the exploration of possible connections.^{56,57}

REFERENCES

- Barnard, R. and T. Horgan. (2006). "Truth as Mediated Correspondence," *The Monist* 89: 31–50.
- Barnard, R. and T. Horgan. (2013). "The Synthetic Unity of Truth," In Pedersen and Wright (2013a): 180–96.
- Beall, Jc. (2000). "On Mixed Inferences and Pluralism about Truth Predicates," *Philosophical Quarterly* 50: 380–2.
- Beall, Jc. (2013). "Deflated Truth Pluralism," in Pedersen and Wright (2013a): 323–38.
- Beall, Jc. and G. Restall. (2006). *Logical Pluralism*. Oxford: Oxford University Press.
- Blackburn, S. (1998). "Wittgenstein, Wright, Rorty and Minimalism," *Mind* 107: 157–81.
- Blackburn, S. (2013). "Deflationism, Pluralism, Expressivism, Pragmatism," in Pedersen and Wright (2013a): 263–77.
- Cohnitz, D. et al. (eds.). (2014). "Monism, Pluralism and Relativism: New Essays on the Status of Logic," *Erkenntnis* 79 (special issue).
- Coliva, A. and N. J. L. L. Pedersen (eds.) (2017). *Epistemic Pluralism*. Basingstoke: Palgrave Macmillan.
- Cook, R. T. (2010). "Let a Thousand Flowers Bloom: A Tour of Logical Pluralism," *Philosophy Compass* 5: 492–504.
- Cook, R. T. (2011). "Alethic Pluralism, Generic Truth and Mixed Conjunctions," *Philosophical Quarterly* 61: 624–9.
- Cotnoir, A. (2009). "Generic Truth and Mixed Conjunctions: Some Alternatives," *Analysis* 63: 473–9.
- Cotnoir, A. (2013a). "Pluralism and Paradox," in Pedersen and Wright (2013a): 339–50.
- Cotnoir, A. (2013b). "Validity for Strong Pluralists," *Philosophy and Phenomenological Research* 86: 563–79.
- Cotnoir, A. and D. Edwards. (2015). "From Alethic Pluralism to Ontological Pluralism. . . and Back!" *Journal of Philosophy*, 112: 113–40.
- David, M. (2013). "Lynch's Functionalist Theory of Truth," in Pedersen and Wright (2013a): 42–68.
- Devlin, J. (2003). "An Argument for an Error Theory of Truth," *Philosophical Perspectives* 17: 52–82.
- Dodd, J. (2002). "Recent Work on Truth," *Philosophical Books* 43: 279–91.
- Dodd, J. (2013). "Deflationism Trumps Pluralism!" in Pedersen and Wright (2013a): 298–322.
- Dummett, M. (1978). "Truth," in M. Dummett, *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 1–24.
- Edwards, D. (2008). "How to Solve the Problem of Mixed Conjunctions," *Analysis* 68: 143–9.
- Edwards, D. (2009). "Truth-Conditions and the Nature of Truth: Re-Solving Mixed Conjunctions," *Analysis* 69: 684–8.

⁵⁶ For more on epistemic pluralism, see Pedersen (2017) and the papers in Coliva and Pedersen (2017).

⁵⁷ Nikolaj J. L. L. Pedersen's work was supported in part by the Yonsei University Research Fund of 2013 (grant no. 2013-22-170) and in part by grant no. 2013S1A2A2035514 from the National Research Foundation of Korea.

- Edwards, D. (2011a). "Naturalness, Representation, and the Metaphysics of Truth," *European Journal of Philosophy* 21: 384–401.
- Edwards, D. (2011b). "Simplifying Alethic Pluralism," *Southern Journal of Philosophy* 49: 28–48.
- Edwards, D. (2012). "On Alethic Disjunctivism," *Dialectica* 66: 200–14.
- Edwards, D. (2013). "Truth, Winning, and Simple Determination Pluralism," in Pedersen and Wright (2013a): 279–91.
- Eklund, M. (2014). Review of Nikolaj J. L. L. Pedersen and Cory D. Wright (eds.), *Truth and Pluralism: Current Debates*. *Notre Dame Philosophical Reviews* 2014.10.14, <https://ndpr.nd.edu/news/truth-and-pluralism-current-debates/>
- Engel, P. (2013). "Alethic Functionalism and the Norm of Belief," in Pedersen and Wright (2013a): 69–86.
- Ferrari, F., S. Moruzzi, and N. J. L. L. Pedersen. (Ms). *True Pluralism: The Ways of Truth*. Book manuscript.
- Field, H. (2009). "Pluralism in Logic," *Review of Symbolic Logic* 2: 342–59.
- Hjortland, O. (2013). "Logical Pluralism, Meaning-Variance, and Verbal Disputes," *Australasian Journal of Philosophy* 91: 355–73.
- Horgan, T. (2001). "Contextual Semantics and Metaphysical Realism: Truth as Indirect Correspondence," in M. Lynch (ed.), *The Nature of Truth: Classical and Contemporary Readings*. Cambridge, MA: MIT Press, 67–95.
- Horton, M. and T. Poston. (2012). "Functionalism about Truth and the Metaphysics of Reduction," *Acta Analytica* 27: 13–27.
- Horwich, P. (1990). *Truth*. Oxford: Blackwell.
- Jackson, F. (1994). "Realism, Truth and Truth-aptness," *Philosophical Books* 35: 162–9.
- Jackson, F. (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Keefe, S. (2014). "What Logical Pluralism Cannot Be," *Synthese* 191: 1375–90.
- Kim, S. and N. J. L. L. Pedersen. (2018). "Strong Alethic Pluralism," to appear in N. Kellen, N. L. L. Pedersen, and J. Wyatt (eds.), *Pluralisms in Truth and Logic*. Basingstoke: Palgrave Macmillan.
- Kim, S. and N. J. L. L. Pedersen. (Ms). "The Return of the Many: A Critical Appraisal of Moderate Truth Pluralism through Metaphysics. Article manuscript.
- Kölbel, M. (2008). "'True' as Ambiguous," *Philosophy and Phenomenological Research* 77: 359–84.
- Kölbel, M. (2013). "Should We Be Pluralists about Truth?" in Pedersen and Wright (2013a): 278–97.
- Lewis, D. (1979). "How to Define Theoretical Terms," *Journal of Philosophy* 67: 427–46.
- Lynch, M. (2000). "Alethic Pluralism and the Functionalist Theory of Truth," *Acta Analytica* 15: 195–214.
- Lynch, M. (2001). "A Functionalist Theory of Truth," in M. Lynch (ed.), *The Nature of Truth: Classic and Contemporary Perspectives*. Cambridge, MA: MIT Press, 723–49.
- Lynch, M. (2004a). *True to Life: Why Truth Matters*. Cambridge, MA: MIT Press.
- Lynch, M. (2004b). "Truth and Multiple Realizability," *Australasian Journal of Philosophy* 82: 384–408.
- Lynch, M. (2005a). "Functionalism and our Folk Theory of Truth: Reply to Cory Wright," *Synthese* 145: 29–43.
- Lynch, M. (2005b). "Précis to *True to Life*, and Replies to Commentators," *Philosophical Books* 46: 289–91, 331–42.
- Lynch, M. (2006). "ReWrighting Pluralism," *The Monist* 89: 63–84.
- Lynch, M. (2008). "Alethic Pluralism, Logical Consequence and the Universality of Reason," *Midwest Studies in Philosophy* 32: 122–40.

- Lynch, M. (2009). *Truth as One and Many*. Oxford: Oxford University Press.
- Lynch, M. (2013). "Three Questions for Truth Pluralism," in Pedersen and Wright (2013a): 21–41.
- McDaniel, K. (2009). "Ways of Being," in D. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics*. Oxford: Oxford University Press, 290–319.
- McDaniel, K. (2010a). "A Return to the Analogy of Being," *Philosophy and Phenomenological Research* 81: 688–717.
- McDaniel, K. (2010b). "Being and Almost Nothingness," *Noûs* 44: 628–49.
- McDaniel, K. (2017). *The Fragmentation of Being*. Oxford: Oxford University Press.
- Nulty, T. J. (2010). "The Metaphysics of Mixed Inferences: Problems with Functionalism Accounts of Alethic Pluralism," *Metaphysica* 11: 153–62.
- Pedersen, N. J. (2006). "What Can the Problem of Mixed Inferences Teach Us about Alethic Pluralism?" *The Monist* 89: 103–17.
- Pedersen, N. J. (2010). "Stabilizing Alethic Pluralism," *Philosophical Quarterly* 60: 92–108.
- Pedersen, N. J. L. L. (2012a). "True Alethic Functionalism?" *International Journal of Philosophical Studies* 20: 125–33.
- Pedersen, N. J. L. L. (2012b). "Recent Work on Alethic Pluralism," *Analysis* 72: 588–607.
- Pedersen, N. J. L. L. (2014). "Pluralism \times 3: Truth, Logic, Metaphysics," *Erkenntnis* 79: 259–77.
- Pedersen, N. J. L. L. (2017). "Pure epistemic pluralism," in Coliva and Pedersen (2017): 47–92.
- Pedersen, N. J. L. L. (Ms-a). "Grounding Manifestation Functionalism." Article manuscript.
- Pedersen, N. J. L. L. (Ms-b). "Grounding Determination Pluralism." Article manuscript.
- Pedersen, N. J. L. L. (Ms-c). "Truth Pluralism and the Structure of Normativity." Article manuscript.
- Pedersen, N. J. L. L. (Ms-d). "Mixing and Matching: A Strongly Pluralist Account of Mixed Discourse." Article Manuscript.
- Pedersen, N. J. L. L. (Ms-e). "On Moderate Pluralism about Truth and Logic: Truth and Logic as One, Quasi-Truth and Quasi-Logic as Many." Article manuscript.
- Pedersen, N. J. L. L. and D. Edwards (2011). Truth as One(s) and Many: On Lynch's Alethic Functionalism," *Analytic Philosophy* 52: 213–30.
- Pedersen, N. J. L. L. and C. D. Wright (2012). "Pluralist Theories of Truth," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), <http://plato.stanford.edu/archives/spr2012/entries/truth-pluralist/>
- Pedersen, N. J. L. L. and C. D. Wright (eds.). (2013a). *Truth and Pluralism: Current Debates*. New York: Oxford University Press.
- Pedersen, N. J. L. L. and C. D. Wright (2013b). "Pluralism about Truth as Alethic Disjunctivism," in Pedersen and Wright (2013a): 87–112.
- Pettit, P. (1996). "Realism and Truth: A Comment on Crispin Wright's *Truth and Objectivity*," *Philosophy and Phenomenological Research* 56: 883–90.
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: MIT Press.
- Read, S. (2006). "Monism: The One True Logic," in D. D. Vidi and T. Kenyon (eds.): *A Logical Approach to Philosophy: Essays in Memory of Graham Solomon*. Dordrecht: Springer, 193–209.
- Rosen, G. (2010). "Metaphysical Dependence: Grounding and Reduction," in B. Hale and A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press, 109–135.
- Russell, G. (2008). "One True Logic?" *Journal of Philosophical Logic* 37: 593–611.
- Sainsbury, M. (1996). "Crispin Wright: Truth and Objectivity," *Philosophy and Phenomenological Research* 56: 899–904.
- Shapiro, S. (2009). "Review of Michael P. Lynch, *Truth as One and Many*," *Notre Dame Philosophical Reviews*. 2009.09.22, <https://ndpr.nd.edu/news/truth-as-one-and-many/>
- Shapiro, S. (2011). "Truth, Function and Paradox," *Analysis* 71: 38–44.

- Shapiro, S. (2014). *Varieties of Logic*. Oxford: Oxford University Press.
- Sher, G. (1998). "On the Possibility of a Substantive Theory of Truth," *Synthese* 117: 133–72.
- Sher, G. (2004). "In Search of a Substantive Theory of Truth," *Journal of Philosophy* 101: 5–36.
- Sher, G. (2005). "Functional Pluralism," *Philosophical Books* 46: 311–30.
- Sher, G. (2013). "Forms of Correspondence: The Intricate Route from Thought to Reality," in Pedersen and Wright (2013a): 157–79.
- Sher, G. (2015). "Truth as Composite Correspondence," in T. Achourioti, H. Galinon, J. Martínez Fernández, and K. Fujimoto (eds.), *Unifying Philosophy of Truth*. Dordrecht: Springer, 191–210.
- Sher, G. (2016). *Epistemic Friction: An Essay on Knowledge, Truth, and Logic*. Oxford: Oxford University Press.
- Spencer, J. (2012). "Ways of Being," *Philosophy Compass* 7: 910–18.
- Tappolet, C. (1997). "Mixed Inferences: A Problem for Pluralism about Truth Predicates," *Analysis* 57: 209–10.
- Tappolet, C. (2000). "Truth Pluralism and Many-Valued Logics: A Reply to Beall," *Philosophical Quarterly* 50: 382–5.
- Tappolet, C. (2010). "Review of Michael P. Lynch, *Truth as One and Many*," *Mind* 119: 1193–8.
- Turner, J. (2010). "Ontological Pluralism," *Journal of Philosophy* 107: 5–34.
- Turner, J. (2012). "Logic and Ontological Pluralism," *Journal of Philosophical Logic* 41: 419–48.
- Williamson, T. (1994). "A Critical Study of *Truth and Objectivity*," *International Journal of Philosophical Studies* 30: 130–44.
- Williamson, T. (1996). "Unreflective Realism," *Philosophy and Phenomenological Research* 56: 905–9.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.
- Wright, C. (1993). "Davidsonian Meaning-Theory and Assertibility," in C. Wright, *Realism, Meaning and Truth*, 2nd edn. Oxford: Blackwell, 403–32.
- Wright, C. (1994a). "Response to Jackson," *Philosophical Books* 35: 169–75.
- Wright, C. (1994b). "Realism, Pure and Simple?" *International Journal of Philosophical Studies* 30: 147–61.
- Wright, C. (1996). "Précis to *Truth and Objectivity* and Response to Commentators," *Philosophy and Phenomenological Research* 56: 863–8, 911–41.
- Wright, C. (1998a). "Truth: A Traditional Debate Reviewed," *Canadian Journal of Philosophy* 24: 31–74.
- Wright, C. (1998b). "Comrades Against Quietism: Reply to Simon Blackburn on *Truth and Objectivity*," *Mind* 107: 183–203.
- Wright, C. (2001). "Minimalism, Deflationism, Pragmatism, Pluralism," in M. Lynch (ed.), *The Nature of Truth: Classical and Contemporary Readings*. Cambridge, MA: MIT Press, 751–87.
- Wright, C. (2003a). "Comrades Against Quietism," in C. Wright, *Saving the Differences: Essays on Themes from Truth and Objectivity*. Cambridge, MA: Harvard University Press, 125–50.
- Wright, C. (2003b). "Truth in Ethics," in *ibid.*, 183–203.
- Wright, C. (2013). "A Plurality of Pluralisms?" in Pedersen and Wright (2013a): 123–53.
- Wright, C. D. (2005). "On the Functionalization of Pluralist Approaches to Truth," *Synthese* 145: 1–28.
- Wright, C. D. (2010). "Truth, Ramsification, and the Pluralist's Revenge," *Australasian Journal of Philosophy* 88: 265–83.
- Wright, C. D. (2012). "Is Pluralism Inherently Unstable?" *Philosophical Studies* 159: 89–105.
- Wyatt, J. (2013). "Domains, Plural Truth, and Mixed Atomic Propositions," *Philosophical Studies* 166: 225–36.
- Wu, A. (2017). "Logic for Alethic Pluralists," *Journal of Philosophy* 114: 277–302.

PART V

OTHER
APPLICATIONS

CHAPTER 21

THE MORAL TRUTH

MARK SCHROEDER

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights . . .

COMMON sense allows that talk about moral truths makes perfect sense. If you object to the United States' Declaration of Independence's assertion that it is a truth that "all men" are "endowed by their Creator with certain unalienable Rights," you are more likely to object that these rights are not unalienable or that they are not endowed by the Creator, or even that its wording ignores the fact that women have rights too, than that this is not the sort of thing which *could* be a truth. Whether it is a truth or not seems beside the point, anyway; the writers of the Declaration could just as well have written, "We hold it to be self-evident that all men are created equal, and also that it is self-evident that all men are endowed by their Creator with certain unalienable Rights," save only that its cadence would lack some of the poetic resonance of the version which garnered Hancock's signature.

Yet famously, ethical noncognitivists have proclaimed that moral sentences *can't* be true or false—that, like "Hooray!" or "Dammit!," they are not even the *kinds* of thing to be true or false. Noncognitivism is sometimes even *defined* as the view that this is so, but even philosophers who define "noncognitivism" more broadly, as consistent with the idea that moral sentences may be true or false, have believed that they needed to do important philosophical spadework in order to make sense of how moral sentences could be true or false. In this chapter we'll look at the puzzle about moral truth as it is faced by early noncognitivists and by metaethical expressivists, the early noncognitivists' contemporary cousins. We'll look at what it would take for expressivists to "earn the right" to talk about moral truths at all, and in particular at what it would take for them to earn the right to claim that moral truths behave in the ways that we should expect—including that meaningful moral sentences which lack presuppositions *are* true or false, and that

classically valid arguments are truth-preserving. And finally, we'll look at whether the assumptions about truth which expressivists need in order to do these things have any independent plausibility or merit.

21.1 NONCOGNITIVISM AND TRUTH

To understand what noncognitivism is, it is useful to start with Ayer's *emotivist* theory from chapter 6 of *Language, Truth and Logic*. Ayer's was not the first noncognitivist theory, and not even the first noncognitivist theory in English-speaking philosophy;¹ but it was a relatively early noncognitivist theory, and Ayer's explanation of its commitments is still fresh and illuminating, and is a good point of departure for understanding later noncognitivist views.

Ayer explained his theory by comparing the meanings of sentences involving "wrong" to counterpart sentences which don't involve "wrong." He compared 1 to 2, and 3 to 4:

- 1 You acted wrongly in stealing that money.
- 2 You stole that money.
- 3 Stealing money is wrong.
- 4 Stealing money.

Each of these pairs, Ayer alleged, has the same *significance*—that feature of sentences which must pass the verifiability condition, according to Ayer's verificationist theory of meaning.

So the main feature of Ayer's view was that whatever kind of meaning "wrong" has, when we add it to "you stole that money" to get "you acted wrongly in stealing that money," the significance is not affected. And similarly, when we add it to "stealing money" to get "stealing money is wrong," its significance is not affected, either. Intuitively, this is because neither of these sentences is about something more—wrongness—than their counterparts. They are just different—emotionally loaded—ways of saying the same thing. Ayer also suggests that we could perform the same function just as well with a special sort of exclamation mark—instead of saying "stealing money is wrong," we could instead just say, "stealing money!!!" in a special tone of voice, or with especially thick exclamation marks. The word "wrong" doesn't affect *what* we say; it only affects *how* we say it.

Since significance is the aspect of meaning which is subject to Ayer's verificationism, Ayer is committed to holding that since sentence 2 can be verified, sentence 1 can be verified as well—under the very same conditions. Since to verify a sentence is

¹ Axel Hägerström is usually credited with providing the first clear statement of emotivism, followed years later by Ogden and Richards (1923), which is usually cited as the first clear statement of emotivism in English-speaking philosophy. *Language, Truth and Logic* was published in 1936.

presumably to verify that it is true, that means that Ayer's theory predicts that sentence 1 *can* be true—but only under the very same conditions that sentence 2 is true. In contrast to sentence 1, however, sentence 3 would be true only under the same conditions that sentence 4 would be true—but 4 isn't the kind of sentence to be true at all, for it fails to express a complete thought. So Ayer's view seems to predict that sentence 3 can't be true at all—and neither can it be false (since sentence 4 can't be false, either).

These are highly counterintuitive implications. It is counterintuitive that "stealing money is wrong" cannot be true not only because we are ordinarily quite comfortable calling it true (if we agree with it—and calling it false if we disagree with it), but also because—as some of Ayer's earliest critics pointed out—arguments involving "stealing money is wrong" as a premise or conclusion are sometimes valid, but an argument can be valid only if the truth of its premises guarantees the truth of its conclusion. If valid arguments need to be truth-preserving, then it is particularly problematic for it to turn out that "stealing is wrong" cannot be true, because that means that it can't turn out to be the conclusion of a valid argument.

Even though Ayer's view makes room for the thesis that sentence 1 *can* be true, moreover, it seems to say the wrong things about *when* sentence 1 is true. This is so not only because it makes perfect sense for someone (think: Robin Hood) to think that 2 is true but that 1 is not true (if you are having trouble with this one, substitute "giving to charity" or "saving lives" for "stealing money"). It is also so because since arguments are valid when the truth of their premises guarantees the truth of their conclusion, Ayer's view makes the wrong predictions about which arguments are valid. For example, it predicts the following argument to be valid:

- P1 You stole that money.
- P2 If you acted wrongly in stealing that money, then you ought to give it back.
- C You ought to give it back.

But this argument is patently not valid (again, if you balk at this, substitute "saving lives" for "stealing money"). So for both kinds of sentence, Ayer's view makes highly counterintuitive predictions.

Now, because it seemed to be such a clear consequence of Ayer's theory that sentences like "stealing money is wrong" cannot be true, philosophers have often *defined* "noncognitivism" to mean the view that moral sentences cannot be true or false. Clearly, in light of the obvious predictions of Ayer's view about sentence 1, this formulation can't be right even about Ayer. But something roughly in its ballpark was for a long time taken to be definitive of a family of theories which, like Ayer's, held moral words like "wrong" to make a different kind of contribution to the meaning of sentences than other kinds of words—more like the kind of contribution made by exclamation marks or the addition of "damn" than like the addition of "square" or "red" or "ten" or "electron." The view, very roughly, is that adding moral words to a sentence doesn't allow you to *say* anything different; what it allows you to do is to *do* something different with what you say.

21.2 THE DEFLATIONIST MOVE

More recently, however, philosophers who are sympathetic to Ayer's idea that moral words make a different kind of contribution to the meaning of sentences than do words like "square" or "electron" have observed that it seems in principle possible to hold this view while still saying that moral sentences can be true or false. After all, they have suggested, Ayer's basic idea wasn't that moral sentences can't be true or false—that was a consequence which he derived from an underlying picture about what the meaning of moral sentences is like. But maybe that consequence *doesn't* really follow. In that case, we could adopt Ayer's underlying view about what the meanings of moral sentences are like, but deny his view that moral sentences can't be true or false. These philosophers either rejected the label "noncognitivism," which was associated with the definition, "moral sentences can't be true or false," or sought to reclaim it, by arguing that it was better associated with the underlying ideas about meaning which had seemed to motivate the conclusions about truth.

The main idea of these philosophers was to adopt ideas from the *deflationist* theory of truth. Now, the core ideas of deflationism can be and have been characterized in a number of closely related ways, but for our purposes a simple and not inaccurate way of understanding the basic idea is that the word "true" is used to agree, and the word "false" is used to disagree—and that once you have a theory which can explain those two things, there is not really anything more to truth than that. Adopting this deflationist idea, the reclaimers argued that if you think that stealing money is wrong, then you agree with "stealing money is wrong." Since there is nothing more to calling it "true" than agreeing with it, correct use of the word "true" therefore permits—indeed, requires—calling it true. So well-brought-up philosophers like us, who *do* think that stealing money is wrong, should agree that it is true that stealing money is wrong. Consequently, we should agree that it can be true that stealing money is wrong. So we should admit that moral sentences can be true—and a similar argument shows that we should admit that moral sentences can be false.

According to these noncognitivist reclaimers of moral truth, their dispute with cognitivist theories of the meaning of moral sentences is less like the dispute between theists and atheists, and more like the dispute over Malebranchean *occasionalism*. According to Malebranchean occasionalism, not only does God exist, but He is the proximate cause of everything that happens. Malebranchean occasionalists aren't just theists; they are theists who claim that God does a vast amount of work. One way of disagreeing with Malebranchean occasionalism is to be an atheist, and deny that God exists at all. But another way to disagree with Malebranchean occasionalism is to believe in God, but simply hold that He does less work. The noncognitivist reclaimers of moral truth say that their dispute with cognitivist theories isn't about whether there are moral truths, but about how much *work* moral truth needs to do. Cognitivist theorists are like Malebranchean occasionalists in offering truth-conditional theories of the

meaning of moral sentences, using truth to explain meaning. One way of disagreeing with cognitivism—as exemplified by some of the earliest noncognitivist theories—is like atheism: it is to deny that there even are any moral truths. But another, equally good, way of disagreeing with cognitivism, is to believe in moral truths, but simply to hold that they do less work. This is the path of the noncognitivist reclaimers of moral truth.

The idea that noncognitivists can say all of these things has now become commonplace in moral philosophy. But it leaves several challenges remaining. Here I'd like to catalogue three, and then explore in the remainder of this article how the main contemporary cousin of Ayer's noncognitivism, *expressivism*, addresses those three challenges.

The first challenge is that it is too quick to say that it follows from the basic deflationist idea that if you are willing to accept or assert a sentence, then you should be willing to accept or assert that it is true. This is certainly not true, for example, of some of the sentences to which early noncognitivists like Ayer analogized moral sentences. For example, consider the sentence, "Boo, stealing money!," or "stealing money!!!," with its specially thick exclamation marks. These are not the kind of sentence that can be true or false. Even if you are willing to assert them, you need not be willing to agree that they are true, because they are not the right *kind* of sentence to be true. But Ayer explicitly endorsed the view that "stealing money is wrong" has exactly the same kind of meaning as "Boo, stealing money!" or as "stealing money!!!" So there is a serious challenge to explaining why we should say that the former sentence can be true, but the latter can't. This challenge is I think quite a bit more difficult than some people have made it out to be; I won't have very much *new* to say about it in this chapter, but I will explain how it can be seen as playing a role in motivating expressivism.

The second challenge that I want to explore is that there is something unsatisfying about the initial explanation that we have received so far of why moral sentences can be true or false. The explanation given so far tries to show that consistency requires admitting that it is true that stealing money is wrong, if you think that stealing is wrong. So since we well-brought-up people think that stealing money *is* wrong, we are committed, on pain of inconsistency, to admitting that it is true that stealing money is wrong—and hence to admitting that there is a truth about whether stealing money is wrong. The reason why this is so far unsatisfactory is that it is consistent with all of this that since it is *not* true that stealing money is wrong, we should not think that stealing money is wrong. So far as this goes, someone who has no view about whether stealing money is wrong may perfectly well deny that there is or could be any truth of the matter about whether stealing money is wrong, and not be erring in any way. Another way of putting this worry is that, so far as the explanation we've seen goes, the claim that there is a truth about whether stealing money is wrong seems to be, at best, itself simply just another *moral* truth. It is something we are committed to since, being well-brought-up, we think that stealing money is wrong. The challenge is therefore to explain why even a moral agnostic would be mistaken to deny that there is a truth of the matter about whether stealing money is wrong. I think this is an easier challenge than the first, but I have never seen it raised or its answer spelled out clearly, so I will do that later in this chapter.

The third challenge I'm going to focus on is the second of the two kinds of counterintuitive consequence of Ayer's view that sentences 1 and 3 had the same truth-values as 2 and 4. The first counterintuitive consequence was that it didn't seem to match what we ordinarily say—and we've already seen how the deflationist can dismiss that consequence. If we will say that "stealing money is wrong" is true just in case we agree that stealing money is wrong, and that "you acted wrongly in stealing that money" is true just in case we agree that you acted wrongly in stealing that money, then this gets the intuitively right results about when we say these things. But that leaves open whether we deal with the second counterintuitive consequence isolated earlier—namely, that Ayer's view got the wrong results about which arguments are valid. The third challenge is to show not only that arguments with moral premises or conclusions are not *precluded* from being valid by lacking truth values or having the intuitively wrong truth values, but that a noncognitivist view can actually get the *right* results about which arguments are valid—in the sense of being truth-preserving. Simon Blackburn calls this the project of "earning the right to truth," and I'll be illustrating in detail how noncognitivists aspire to do this, by walking through the way that it can be done by expressivism.

21.3 THE FIRST CHALLENGE AND TRUTH-APTNESS

Our first challenge is the challenge of drawing a line around *which* sentences we are supposed to admit are true, if we are willing to endorse them in the first place. The noncognitivist wants to say that "stealing money is wrong" is among those to which this deflationist idea applies, but that "Boo, stealing money!" is not. The mere fact that you are willing to go around saying, "stealing money is wrong," requires you to be willing to assent to "the sentence, 'stealing money is wrong,' is true" as well, but the mere fact that you are willing to go around saying "Boo, stealing money!" does not require you to be willing to assent to "the sentence, 'Boo, stealing money!' is true." Since the latter sort of sentence *can't* be true, the challenge is to draw the line around the sort of sentences to which the deflationist idea applies in a way that includes "stealing money is wrong" but does not include "Boo, stealing money!"

Another name for this challenge is the problem of articulating a satisfactory theory of *truth-aptness*. We may say that a sentence is *truth-apt* if it is the sort of sentence that, when you are willing to say it, you should be willing to allow is true. Now, deflationism about truth alone can't establish that moral sentences can be true or false—only deflationism together with the view that moral sentences are truth-apt can establish that those of us who accept some moral sentences, must admit that moral sentences can be true. Now every version of deflationism needs a theory of truth-aptness—a story about which sentences "P" can be replaced with "it is true that P" or with "P" is true." But noncognitivists face a special challenge in outlining a

theory of truth-aptness, because a noncognitivist theory of truth-aptness must classify moral sentences like “stealing is wrong” differently from sentences like “Boo, stealing money!” and “Dammit—not stealing money!,” even though the central idea of noncognitivism, as we’ve seen, is that the former sentences have the very same kind of meaning as the latter sentences. The fact that noncognitivists think these sentences have the very same kind of meaning—perhaps even the *same* meaning—is the principal challenge to any theory of truth-aptness being able to classify moral sentences as truth-apt, but not the others.

Now, there is a very simple way in which some philosophers have believed that this trick could be accomplished. According to these philosophers, the distinction between truth-apt and non-truth-apt sentences is not a semantic distinction at all. It is simply a *syntactic* distinction. The test for whether a sentence “P” is truth-apt or not, they have said, is simply whether “it is true that P” is grammatical (in the sense of being syntactically well-formed) or not. Since on this view the distinction between truth-apt and non-truth-apt sentences is not a semantic one at all, but a purely syntactic one, it runs into no difficulty at all with the hypothesis that “stealing money is wrong” and “Boo, stealing money!” have *exactly* the same meaning, but the former is true and the latter is not.

This is because despite the fact that these two sentences are semantically identical, they are syntactically quite different. Indeed, the sentence “it is true that stealing money is wrong” is grammatical, but the sentence “it is true that Boo, stealing money!” is not. So this syntactic theory of truth-aptness not only tells us which sentences are truth-apt, and not only classifies this pair of sentences in the way that the noncognitivist hopes for, but does so in a way that is compatible with the very strongest sort of noncognitivism—with the view that these two sentences have exactly the same meaning.

Unfortunately, there are problems with the idea that truth-aptness is a merely syntactic property, rather than having any basis in semantics. This idea fares pretty well with how we actually classify existing sentences of English, sure enough, but it also predicts that all that we need to do in order to make a non-truth-apt sentence truth-apt is to change its grammar. Dreier (1996) tested this prediction by introducing a new predicate, “hiyo.” According to Dreier, “Bob is hiyo” is used in very much the same way as “hey, Bob” is used—to accost Bob. But according to Dreier, “hiyo” is grammatically an adjective, so “Bob is hiyo” is an ordinary indicative sentence, and consequently “it is true that Bob is hiyo” is syntactically well-formed. Consequently, by the syntactic criterion, “Bob is hiyo” is truth-apt, and so someone who is willing to accost Bob should be willing to admit that “Bob is hiyo” is true. Dreier calls this view about “hiyo” *accostivism*.

It is far from clear, however, whether this should really be so. In the case of moral sentences, we certainly do ordinarily say that they can be true or false. And they are certainly indicative. But what is at issue here is precisely whether it is their indicative syntax alone which suffices to guarantee that they can be true or false. And it isn’t clear whether “Bob is hiyo” really should be said to be true or false, simply because we are willing or unwilling to accost Bob.

Dreier sharpens the worry by noting that sentences like “if Bob is hiyo, then I’m out of here” and “everyone who isn’t hiyo is boring” are *also* grammatically well-formed,

if “hiyo” is an ordinary predicate—and indeed, are themselves truth-apt, by the syntactic criterion for truth-aptness. Moreover, since these sentences are formed truth-functionally on the basis of their parts (assuming for the sake of argument the material conditional treatment of “if . . . then” for the former case), that tells us under what conditions these sentences *are* true. But as Dreier notes, even once we understand that “Bob is hiyo” is used to accost Bob, so that we understand how to use *that* sentence, we *still* don’t understand what sentences like “everyone who isn’t hiyo is boring” mean, or know how to use them. A theory which tells us that it is this easy to confer a meaning on these sentences makes too much out of syntax.

Dreier’s discussion pushes the noncognitivist to explain the difference between “stealing is wrong” and “Boo, stealing!” on *semantic* grounds. The bare syntactic criterion on truth-aptness makes it too easy for sentences to turn out to be true or false, and consequently too easy to generate putatively meaningful complex sentences which are patently not meaningful at all. What noncognitivists need, in order to avoid the result that “hiyo” is a perfectly intelligible predicate, is some kind of semantic constraint which, intuitively, goes hand-in-hand with indicative syntax in natural languages, but is violated when Dreier tries to simply stipulate indicative syntax for sentences involving “hiyo.”

The best existing answer to Dreier’s challenge that I know of comes from Allan Gibbard. Gibbard’s idea is closely connected to my original characterization of deflationism as the idea that “true” is used to agree, and “false” is used to disagree. The rough idea is that some sentences are not used to do things that it is possible to agree or disagree with—and that such sentences are not truth-apt. According to Gibbard, it is not possible to agree or disagree with an accosting, and so it is not possible to agree or disagree with “Bob is hiyo.” Consequently, it does not make sense to say that “Bob is hiyo” is true or false, even if you are willing or unwilling to accost Bob—and even though it is syntactically indicative.

In the abstract, it is obscure to talk about agreeing or disagreeing with what a sentence is “used to do.” I used that expression in order to give a sense for Gibbard’s answer to Dreier. But to fully understand Gibbard’s answer to Dreier, we need to understand the *kind* of noncognitivist theory which Gibbard endorses, which is known as *expressivism*. In the next section I’ll explain the basic ideas of expressivism, and then restate Gibbard’s answer to Dreier in that framework. Then in the remainder of the chapter we’ll go on to address the second and third challenges in the context of expressivism.

21.4 CONTEMPORARY NONCOGNITIVISM: EXPRESSIVISM

In my view, it is most promising and helpful to think of expressivism as a kind of *assertability conditional* semantic theory. On this interpretation, the main idea of

expressivism is that just as the phonological rules of a language set *phonological* correctness conditions, and the syntactic rules of a language set *syntactic* correctness conditions (e.g. “colorless green ideas sleep furiously” meets these but “example one this good a is” fails them, even if we can guess what it means), the semantic rules of a language set *semantic* correctness conditions. But on this view, semantic correctness conditions aren’t *truth* conditions. To see why not, imagine that Todd followed the Drudge Report up until the end of October 2008, but then subsequently stopped paying attention to the US presidential election or its outcome, and consequently believes that John McCain was on track to become, and now is, the president of the United States. When asked who is the president of the United States, Todd volunteers, “John McCain is president of the United States.” What Todd has said is not true, and so in asserting this he has made a mistake. But intuitively, Todd’s mistake is not a linguistic one; it is a mistake about US electoral politics.

Examples like this one motivate the view that the semantic correctness condition of “P” is not that P, but rather that the speaker thinks that P. It is because Todd thinks that John McCain is president of the United States that when he says so, he makes no semantic mistake. In general, on this view, a semantic theory needs to associate each sentence “P” of the language with a mental state: what it is to think that P. The semantic correctness condition of uttering “P” is then that the speaker is in that mental state. We say that the state associated by the semantics with a sentence is the state that that sentence *expresses*.

What makes expressivism a way of developing noncognitivism is the observation that within a semantic framework of this kind, it need not turn out that what it is to think that P is always to have an ordinary belief about the world. According to metaethical expressivism, when “P” is an ordinary descriptive sentence like “grass is green,” what it is to think that P is to have an ordinary descriptive belief—in this case, the belief that grass is green. But when “P” is a moral sentence like “stealing money is wrong,” what it is to think that P is not to have an ordinary belief about the wrongness of stealing money at all. Instead, it is to have a negative desire-like attitude toward stealing money—to *disapprove* of stealing money.

Disapproval of stealing money is only an attitude toward stealing money—and not an attitude toward wrongness. So it differs importantly from the belief that grass is green, which is an attitude not just toward grass, but toward its greenness. So metaethical expressivists agree with Ayer that moral words like “wrong” make a different kind of contribution to the meanings of the sentences involving them than do ordinary descriptive words like “green.” Rather than contributing to what accepting such sentences is having an attitude *about*, moral terms like “wrong” affect which *kind* of attitude is involved in accepting sentences involving it.

In the terms of expressivism, Gibbard’s response to Dreier is that truth-apt sentences must express states of mind with which it is possible to agree or disagree. It is possible to agree or disagree with ordinary descriptive beliefs, like the belief that grass is green. You disagree with it for example if you believe that grass is not green. In general, any two beliefs with inconsistent contents “disagree” with one another, in the sense that anyone who had the

first of these beliefs would disagree with anyone who had the second. So ordinary descriptive beliefs are uncontroversially the kind of state of mind with which it is possible to agree or disagree.

Following Stevenson, however, Gibbard holds that ordinary descriptive beliefs are not the only kinds of state of mind with which it is possible to agree or disagree. For example, it is also possible to disagree in intention. If the two of us are making plans about what to do tonight, I may intend that we go to the symphony, since I can't stand movies, and you may intend that we go to the cinema, since you can't stand classical music. If so, then intuitively, we disagree about what to do tonight. Gibbard holds that disapproval of stealing money is also an attitude that it is possible to disagree with. For example, if you disapprove of stealing money, and I disapprove of *not* stealing money, then we seem to be in some sort of disagreement. If this is right, then disapproval is the right sort of attitude to be expressed by truth-apt sentences. In contrast, it is not possible to disagree with a headache. So no sentence which expressed the state of having a headache could be truth-apt. And similarly, it is not possible to disagree with an accosting. So Dreier's stipulations about "hiyo" do not suffice to make "Bob is hiyo" truth-apt.

So now we've encountered the very basic idea of expressivism, and seen how it enables us to preserve some of the basic ideas underlying Ayer's emotivism, but in a way that enables a well-motivated answer to Dreier, and consequently a good explanation of how "stealing money is wrong" could be truth-apt, even though "Boo, stealing money!" and "Bob is hiyo" are not. In section 21.5, I'll explain the main problem facing noncognitivism—including expressivism—which is often known as the *Frege-Geach* problem, and what it has to do with truth. That will set us up to better understand what expressivist attempts to explain validity are like and what an expressivist account of truth would look like, in sections 21.6 and 21.7. Finally, in section 21.8 we'll reap the payoff of this extended discussion of expressivism, by seeing how expressivists can hope to address my second and third challenges. And in section 21.9 we'll examine the undischarged assumptions we needed to make on behalf of expressivists along the way.

21.5 THE FREGE-GEACH PROBLEM

The most famous and pressing problem facing any version of noncognitivism—expressivism included—is that of providing an adequate compositional semantics: of accounting for the meanings of complex sentences, and doing so on the basis of the meanings of their parts. To get an initial sense for this problem, recall Ayer's view that "wrong" does not add anything to the significance of sentences in which it figures, and look at some sentences which are even less congenial to this proposal than the examples that he discusses:

- 5 Everything wrong is forbidden in the Bible.
- 6 Everything is forbidden in the Bible.

- 7 Stealing money is wrong or my parents lied to me.
- 8 Stealing money or my parents lied to me.

Obviously here we have a problem; sentence 5 is not merely a different way of saying sentence 6—on the contrary, most people who would assent to 5 would deny 6. Even worse, sentence 8 isn't even well-formed; you can't take the disjunction of a gerundival phrase and a sentence. Certainly if Ayer's view is committed to thinking that you can, then he has his work cut out for him in explaining to us what it means.

The problem is that it is all well and good for Ayer to tell us what simple sentences involving "wrong" mean, by comparing "stealing money is wrong" to "stealing money!!!" But an adequate theory of meaning must not only tell us the meanings of simple sentences; it must also provide us with a way of determining the meanings of complex sentences. But as the examples just considered illustrate, Ayer did not get very far in considering the implications of his view for complex sentences, and as the example of sentence 7 illustrates, the ordinary compositional rules which work for other sorts of semantic theory do not look like they are capable of doing the sort of work which Ayer's view requires of them.

In the early 1960s, Peter Geach pressed this problem against noncognitivism as he interpreted it.² What Geach did was to collect pieces of evidence that moral sentences have the same meaning, even when they appear as part of complex sentences. For example, compare "stealing money is wrong" to "is it the case that stealing money is wrong?" The former of these sentences is an *answer* to the latter, and that is so precisely because the words "stealing money is wrong" mean the same thing in both sentences—if any of those words were ambiguous across the two sentences, like the ambiguity in "bank," then the former sentence would not be the answer to the latter. Another example Geach considered was the connection between "stealing money is wrong" and "it is not the case that stealing money is wrong." Again, Geach argued, "stealing money is wrong" means the same thing as it appears in both sentences—and the evidence for that is that the two sentences are *inconsistent*.

The most famous example used by Geach was intended to show precisely the same thing, about a different pair of sentences. He considered sentences like "stealing money is wrong," "if stealing money is wrong, then embezzlement is wrong," and "embezzlement is wrong." He noted that the first two sentences provide a valid argument for the third, and again contended that this is because "stealing money is wrong" has the same meaning in both places in which it appears, and that "embezzlement is wrong" has the same meaning in both places in which it appears.

Geach made a big deal out of his arguments that the words "stealing money is wrong" have the very same meaning when they appear in complex sentences as they do when they appear unembedded, because he believed that this was inconsistent with

² John Searle (1962) independently pressed a very similar objection, but for Searle the dialectic was slightly more complicated; see Schroeder (2008c) for discussion.

noncognitivism. As Geach understood noncognitivism, what gives moral sentences the meaning that they have is what they are used to *do*. For example, Ayer seemed to believe that what gives “wrong” its meaning is the emotion it is used to express; similarly Stevenson seemed to believe that what gives “wrong” its meaning is the emotion it is used to encourage one’s audience to have; and Hare seemed to believe that what gives “wrong” its meaning is that it is used to condemn. But, Geach noted, the question, “is stealing money wrong?” is not used to express the same negative emotion as “stealing money is wrong,” nor is it used to encourage one’s audience to have that same emotion, nor is it used to condemn stealing. So, Geach concluded, noncognitivists are committed to denying an obvious and important truth: namely, that “wrong” means the same thing when it appears in questions as when it appears in indicative sentences. And he made similar points for negations and conditionals.

The consensus response to the problem as posed by Geach is that moral sentences like “stealing money is wrong” don’t have to be used to do the very same thing in complex sentences as they are used to do when unembedded, in order to count as having the same meaning. All that has to happen is that the meaning of the complex sentences in which they figure is a *function* of the meanings of the constituent sentences. This is the answer which Hare gave to Geach (and to Searle, who offered a similar argument against noncognitivism) in 1970, and it has informed research on the Frege-Geach problem ever since.

In light of Hare’s response to Geach, we can think of the problem facing noncognitivism (including expressivism) as being that of providing compositional rules which determine the meanings of questions, negations, and conditionals (to say nothing of other sorts of complex sentences) as a function of the meanings of their parts—and all within a noncognitivist semantic framework. Within an expressivist theory, what this means is that an account of the meaning of “not” must provide a function from mental states to mental states, which associates the mental state expressed by any sentence “P” with the mental state expressed by its negation, “~P.” Similarly, an expressivist account of the meaning of “if . . . then” must provide a function from pairs of mental states to mental states, which associates the mental states expressed by any pair of sentences “P” and “Q” with the mental state expressed by the conditional formed from them, “P→Q.” And so on for other complex-sentence-forming constructions.

The central constraint on this enterprise derives from the fact that not just any functions will do; these functions must assign the *right* mental state to the complex sentences, in a way that allows us to predict and explain the semantic properties of complex sentences. And Geach’s argument provides us with an initial list of just what those semantic properties are: yes-no questions should turn out to be answered by their indicative counterparts, negations should turn out to be inconsistent with the sentences they negate, and conditionals should turn out to license *modus ponens*. No compositional semantic theory will be adequate unless it can explain these basic things about each of these constructions.

It is important to understand that the Frege-Geach problem is not fundamentally a problem about truth. It is a fully general problem about whether a non-standard

semantic theory like expressivism can account for the compositional features of natural languages. But truth does come into the picture, because some of the important semantic properties of sentences formed by the connectives of propositional logic (for example) *do* seem to have to do with truth. For example, the most important semantic property of “not” seems to be that negations are inconsistent with the sentences that they negate—i.e. that “stealing money is not wrong” is inconsistent with “stealing money is wrong.” But inconsistency would ordinarily be defined in terms of truth—two sentences are inconsistent just in case they can’t both be true. Similarly, the most important semantic property of conditionals is usually taken to be that they validate *modus ponens*. But validity would ordinarily be defined in terms of truth—an argument is valid just in case the truth of its premises guarantees the truth of its conclusion. And so these observations get us back to the problem: just how are expressivists going to make sense of moral truth?

The usual expressivist strategy for explaining these things is to divide the semantic properties of sentences up into two classes, a primary class, and a secondary class. The primary class of semantic properties of sentences includes properties about the rational relationships between *accepting* the sentences (more on this directly below); the secondary class of semantic properties of sentences includes all of their other semantic properties, including facts about which sentences they are inconsistent with (in the sense that both can’t be true) and which valid arguments they figure in (in the sense of which preserve truth). Then expressivists divide their theory up into two stages. At the first stage, they propose compositional theories of which mental states are expressed by complex sentences as a function of the mental states expressed by their parts, and try to use those functions in order to explain the primary class of semantic properties of those sentences. At the second stage, they try to *use* the primary semantic properties of sentences in order to predict and explain their secondary semantic properties—including facts about inconsistency and validity.

Much of the attention that has been paid to expressivist theories and in particular to expressivist solutions to the Frege-Geach problem has focused on the first of these two stages—on which mental states the theory says are expressed by complex sentences. In what follows, I will instead focus on the second of these two stages. I will be offering an approach to understanding what happens at this second stage of expressivist theorizing—an approach that I call *commitment theory*. I believe that commitment theory, as I will be developing it here, is neutral between all existing expressivist theories. Studying commitment theory illustrates two things: first, how much expressivists can hope to accomplish—including providing the answers to our second and third challenges—if only they can manage to discharge the first stage of their theory. And second, it shows us exactly what intermediate goal the first stage of expressivist theorizing needs to meet, in order to be able to have these further payoffs.

We can therefore think of commitment theory as trying to reduce to a very manageable size the number of things which expressivists need to be able to explain at the first stage of their theory, and showing how much more can be explained on the basis of those resources. In sections 21.6 and 21.7 I’ll explain how commitment theory works,

and then in section 21.8 we'll apply it in order to answer our second and third challenges about moral truth.

21.6 COMMITMENT THEORY—THE BASICS

Commitment theory assumes that since expressivism assigns each sentence, “P” to a mental state—to what it is to think that P—it should be in principle possible to use general principles from the philosophy of mind in order to explain the rational relationships between these mental states. So the idea is that a good place for expressivists to start, in explaining the semantic properties of a sentence like “stealing money is not wrong,” is to start by explaining the rational relationship between thinking that stealing money is not wrong, and having other sorts of thought. In particular, a good place to start would be by explaining why it is *rationally inconsistent* to think these two things at the same time.

Note that the sense in which it is rationally inconsistent to think that stealing money is wrong and also think that stealing money is not wrong, is not just that “stealing money is wrong” and “stealing money is not wrong” are inconsistent. Even though these are inconsistent, there is no rational inconsistency in supposing that stealing money is wrong and also supposing that stealing money is not wrong, or in wondering whether stealing money is wrong and also wondering whether stealing money is not wrong. So there is more to *rational inconsistency* in the sense at stake, here, than merely a matter of having thoughts which can't both be true. Gibbard (2003) calls the kind of rational inconsistency that we are interested in *disagreement*; the idea is that when you think that stealing money is wrong and someone else thinks that it is not wrong, you disagree with her; similarly to think both at the same time would be to disagree with yourself.

The fact that it is rationally inconsistent to think that stealing money is wrong and to also think that stealing money is not wrong is just a special case of a much more general fact. Whenever an argument is classically valid, it is rationally inconsistent to accept all of its premises and deny its conclusion. The original observation is a special case of this more general principle, because the single-premise argument whose premise and conclusion are both “stealing money is wrong” is classically valid. Let us say that an argument has the *inconsistency property* just in case it is rationally inconsistent to accept its premises and deny its conclusion. One thing that expressivists ought to hope to be able to explain, therefore, is the fact that all classically valid arguments have the inconsistency property.

A related but different property that an argument might have is the property of being *inference-licensing*. Let us say that an argument *licenses inference* just in case someone who accepts its premises is *committed* to accepting its conclusion. Since all theorems of classical logic are the conclusions of classically valid arguments from vacuous premises, it follows that if all classically valid arguments license inference, then everyone is committed to *accepting* all classical theorems, no matter what else they accept. Similarly,

since all classically valid arguments have the inconsistency property, it is rationally inconsistent for anyone to deny any classical theorem, no matter what else they accept.

So long as it is rationally inconsistent to accept and deny one and the same sentence, the former of these two claims entails the latter. But the converse entailment is more controversial. Some philosophers believe that it sometimes makes sense to *reject* a sentence, neither accepting nor denying it, but not merely withholding, or simply waiting to make up one's mind, either. If it makes sense to reject "P," some of these philosophers believe, then it also makes sense to reject " $P \vee \sim P$." On this view, it is always inconsistent to *deny* a classical theorem, but we are not always committed to *accepting* them. So on this view, not all classically valid arguments license inference—in particular, arguments which rely on steps with vacuous premises do not. On this picture, the inferences which are licensed are those whose conclusions are "relevant" to their premises. So the inference-licensing property can come apart, at least in principle, from the inconsistency property.

Rather than evaluate whether there really is such an attitude as rejection, or whether it is ever rational to reject a sentence, I am going to leave this question open, by working with a framework that leaves the possibility of rejection open, but which also tells us what happens if there is no such attitude, or if it is never rational. I will assume, therefore, that there are *at most* three possible attitudes to have toward a sentence, "P"—acceptance, rejection, and denial. I will also assume that denying "P" is the same as accepting " $\sim P$," and (this is the "at most" part) that these are the *only* positive attitudes one can have toward "P" in the sense that someone who does not have one of these attitudes toward P is undecided whether P. And finally, I will assume that each of these three attitudes is rationally inconsistent with either of the other two attitudes toward one and the same sentence. These are the main assumptions of commitment theory.

21.7 COMMITMENT THEORY— THE SUBSTANCE

With those assumptions in hand, we can exploit the notion of a *commitment table*, in order to map out which of those three attitudes a thinker is committed to having toward a sentence, on the basis of the attitudes that she has toward other sentences. For example, consider the following table:

P	$\sim P$
A	D
R	R
D	A

This is the commitment table for " \sim " ("not"). It tells us that someone who accepts "P" is committed to denying " $\sim P$ " (i.e. to accepting " $\sim \sim P$ "), that someone who rejects

“P” is committed to rejecting “ $\sim P$ ” and that someone who denies “P” is committed to accepting “ $\sim P$ ” (in fact, by assumption, these two states are identical).

Don’t be misled into thinking that this is a truth-table; it is not, though it does look like one. The table tells us nothing about the semantic status of either “P” or “ $\sim P$ ”—it only tells us which combinations of views about “P” and “ $\sim P$ ” are rationally consistent, and which are not. A pair of views is rationally consistent if it corresponds to a row on the table; otherwise not. (In deriving commitment tables like this one, we can either take as primitive the notion of commitment, and explain rational inconsistency as being committed to having different attitudes to the same sentence, or we can start with a notion of rational inconsistency, and explain commitment, by saying that states $S_1 \dots S_n$ commit to attitude A to “P” just in case of the three attitudes toward “P,” A is the only one that is not rationally inconsistent with the combination of $S_1 \dots S_n$.)

Similarly, the commitment table for “&” would be:

P	Q	P&Q
A	A	A
A	R	R
A	D	D
R	A	R
R	R	R
R	D	D
D	A	D
D	R	D
D	D	D

Together the commitment tables for “ \sim ” and “&” suffice to derive commitment tables for every sentence of propositional logic. Moreover, facts about these two tables suffice to establish that it is rationally inconsistent to deny any theorem of classical propositional logic. This is because these commitment tables are structurally identical to what are known as the strong Kleene 3-valued truth-tables, and so it is a well-known fact that the commitment tables for all and only classical theorems never contain a “D”—which means that there are no consistent states of mind which involve denying them.³ (If we assume that rejection is not really a rational possibility, then this turns into the result that everyone is committed to *accepting* every classical theorem.)

Moreover, it is a fact about classical logic that whenever $P_1 \dots P_n$; Q is a classically valid argument, $(P_1 \& \dots \& P_n) \supset Q$ is a classical theorem—which means that it is inconsistent to deny. But it is straightforward to prove from our commitment tables that someone who accepts all of $P_1 \dots P_n$ as well as $\sim Q$ is committed to denying $(P_1 \& \dots \& P_n) \supset Q$, given the appropriate definition of “ \supset ” in terms of “&” and “ \sim .” That is, they are committed to denying something that it is rationally inconsistent to deny, if the argument from $P_1 \dots P_n$ to Q is classically valid. So this suffices to establish that

³ For the proof, see e.g. Avron (1991). This result is equivalent to the result that the logic LP has all and only the theorems of classical logic. This is because LP results from the strong Kleene truth-tables under the treatment of both “A” and “R” as “designated.”

classically valid arguments all have the inconsistency property—just on the basis of our two commitment tables.

Similarly, our two tables suffice to establish that all *relevance*-valid arguments have the inference-licensing property, and that if rejection is not rationally possible, then all classically valid arguments have the inference-licensing property. This is because to say that an argument is inference-licensing is to say that accepting its premises commits you to accepting its conclusion—which is to say that the commitment table for the conclusion should have only “A’s, wherever there are “A’s for all of the premises. But it is a well-known fact about the strong Kleene truth-tables that all “relevance-valid” arguments have this feature.⁴ The following table illustrates the case for *modus ponens* (observe that every row with an “A” for both “P” and “P \supset Q” also receives an “A” for “Q”):

P	Q	P \supset Q
A	A	A
A	R	R
A	D	D
R	A	A
R	R	R
R	D	R
D	A	A
D	R	A
D	D	A

Similarly, if we assume that rejection is not rationally possible, then we can simply delete every row containing an “R,” and our commitment tables turn out to be structurally identical to the classical truth-tables, with “A” substituted for “T” and “D” substituted for “F.” Obviously, every classically valid argument preserves “T”—has only “T”’s on its truth-table, wherever “T” is assigned to all of its premises. So similarly, every classically valid argument preserves “A”—has only “A”’s on its commitment table, wherever “A” is assigned to all of its premises. That is, accepting its premises commits you to accepting its conclusion—it licenses inference.

It is a consequence of these observations, that these two commitment tables reduce the challenge facing expressivist accounts of the logical sentential connectives in the *first* stage of expressivist theorizing to the challenge of predicting and explaining these two commitment tables. The idea is to provide a recipe for what it is to accept “ \sim P” on the basis of what it is to accept “P” which predicts and explains why accepting “P” commits to rejecting “ \sim P”—i.e. to accepting “ $\sim\sim$ P” and conversely; and further (if you go in for rejection)

⁴ For our purposes here, we may understand an inference to be relevance-valid only if the conclusion shares some propositional variable with the premises. The basic idea is that the only classically valid arguments for which our commitment tables don’t assign an “A” to the conclusion whenever they assign “A” to all of the premises are ones which rely on classical theorems involving sentences which don’t appear in the premises—e.g. as in the argument from “P” to “ $\sim(Q \& \sim Q)$.” We can think of “ $\sim(Q \& \sim Q)$ ” as an auxiliary premise which, though undeniable, need not be accepted—and hence someone who rejects it need not be committed to “ $\sim(Q \& \sim Q)$,” even though she is committed to “P” and the argument from “P” to “ $\sim(Q \& \sim Q)$ ” is classically valid.

to provide a recipe for what it is to reject “P” on the basis of what it is to accept “P” which explains why rejecting “P” commits to rejecting “ \sim P” and conversely, and is rationally inconsistent both with accepting “P” and with accepting “ \sim P”; and finally, to provide a recipe for what it is to accept “P&Q” on the basis of what it is to accept “P” and to accept “Q” which, together with the other recipes, explains why accepting both “P” and “Q” commits to accepting “P&Q” and conversely, and so on for each of the other properties of the commitment table for “&.” Thus so long as our expressivist account can explain the properties that are captured in the commitment table, then we can use those results (as we have done here) in order to predict and explain at least one important property of all classically valid arguments. (The next step—for section 21.8—will be to appeal to this property of classically valid arguments, in explaining why they are truth-preserving.)⁵

So while none of this, I think, tells us what the first stage of an expressivist semantic theory should look like—none of it tells us what it is to accept complex sentences involving “ \sim ” or “&”—it does set us a reasonable intermediate goal that seems like at least the *kind* of thing that an expressivist account ought to be able to aspire to explain. The idea is, we start with rational relationships between mental states, and then we try to explain other things from there. It is fruitful to start looking for how an expressivist account of truth would work, by looking in the same place: rather than starting from the beginning by asking, “What is it to accept sentences involving ‘true’?”, we want to start mid-way, by trying to characterize some criteria that an adequate answer to the former question might be expected to meet. That, in turn, will help to reduce the task faced by the former question and help us to better understand the constraints that it faces.

So we can think of an expressivist approach to truth in this same framework. In order to avoid issues about whether expressivists can quantify over propositions—including over moral propositions—I’ll focus here on a sentential truth predicate. Now obviously, the commitments of someone who thinks a sentence S is true, depend on what she thinks S means. So the commitment table for “S is true” will need to be sensitive to what you think S means. The relevant part of the table looks like this:

P	S means that P	S is true
A	A	A
R	A	R
D	A	D
A	R	
R	R	
D	R	
A	D	
R	D	
D	D	

⁵ For an illustration of the first stage of an expressivist theory which does exactly this, see Schroeder (2008: chs. 4–8, esp. ch. 8, section 1).

I'm leaving the bottom part of the column for "S is true" blank, largely because it won't matter for our purposes how it is filled in—though it is worth noting that things are a little bit more complicated here, since the section where "S means that P" receives a "D" column will not be a function of the thinker's attitudes with respect to "P" and to "S means that P."

This much of the table, however, is sufficient to establish that when someone accepts "S means that P" she is committed to having the same attitude toward "S is true" and toward "P"—whatever that attitude is. Moreover, since her commitment toward complex sentences formed by the connectives of propositional logic is commitment-functional (as spelled out by the commitment tables for " \sim " and "&") that means that "P" and "S is true" can be substituted everywhere in sentences formed by such connectives, and preserve commitment (for someone who accepts "S means that P"). This is a deeply significant result. It means that "P" and "S is true" are completely intersubstitutable—that any inference substituting one for the other licenses inference, so long as "S means that P" is also accepted.

21.8 APPLICATIONS: THE SECOND AND THIRD CHALLENGES DISPOSED OF

We can use this result about the intersubstitutability of "P" and "S is true" in order to derive many important observations about truth. For example, we can use it to show that for someone who thinks that "stealing is wrong" means that stealing is wrong and "stealing is not wrong" means that stealing is not wrong, "stealing is wrong" is true or 'stealing is not wrong' is true" will have the same commitment status as "stealing is wrong or stealing is not wrong"—i.e. it will be rationally inconsistent to deny it, and everyone will be committed to it, if rejection is not a rational possibility. So as long as someone admits that "stealing is wrong" means that stealing is wrong, on this view she cannot rationally deny that there is a truth of the matter about whether stealing is wrong—even if she herself has no opinion about whether stealing is wrong or not.

This means that on this view, the claim that there is a truth of the matter about whether stealing is wrong is *not* simply a moral truth. It is not simply something that people will say, who either think that stealing is wrong, or who think that stealing is not wrong. It is something that *no one* who thinks that "stealing is wrong" means that stealing is wrong and that "stealing is not wrong" means that stealing is not wrong can consistently deny. The claim that there is a truth of the matter about whether stealing is wrong therefore has a special status due to the meaning of "true"—not just to the fact that we well-brought-up folks in fact think that stealing is wrong. So that is an answer to our second challenge.

Similar arguments can be used to explain why classically valid arguments are truth-preserving, answering our third challenge and completing the response to the

Frege-Geach problem by explaining how conditionals validate *modus ponens*. We already showed above that classically valid arguments have the inconsistency property, and in particular, that when $P_1 \dots P_n; Q$ is a classically valid argument, it is rationally inconsistent to deny $(P_1 \& \dots \& P_n) \supset Q$. But that means that for someone who thinks that “ P_1 ” means that P_1 , ... “ P_n ” means that P_n , and “ Q ” means that Q , “ $((P_1 \text{ is true}) \& \dots \& (P_n \text{ is true})) \supset (Q \text{ is true})$ ” will also be rationally undeniable.

So now we make the transcendental turn: since it is undeniable that if the premises of a classically valid argument are true, then the conclusion is true as well, let’s not deny it. Moreover, since rejecting it commits us to rejecting at least one of its atoms, let’s not reject it, either. Consequently, let us accept it. So since we accept it, let us say it: if the premises are true, then the conclusion is true as well.

This is how expressivists can “earn the right” to truth. The first step is to provide an expressivist account of truth, which tells us which mental states are expressed by sentences involving “true”—i.e. what it is to accept such sentences. We haven’t explored that step here—instead we have formulated an intermediate stage, where we offered conditions of adequacy for what such an account should be able to establish, in the form of commitment tables for “ \sim ,” “ $\&$,” and “true.” The second step is to show how, given the first step, “ s is true” is substitutable everywhere for “ P ,” in the commitments of anyone who thinks that S means that P . The third step is to use that fact in order to show that the claim that if the premises of an argument are true, then the conclusion is true as well, has the same status as the theorems of classical logic. And finally, the fourth step, since the claim that such arguments preserve truth has the same status as the truths of classical logic, to make the transcendental turn, and accept that claim.

21.9 WHAT REMAINS FOR EXPRESSIVIST TRUTH?

Commitment theory doesn’t tell us how expressivists can earn the right to truth; it only tells us how expressivists can *hope* to earn the right to truth. To make good on this hope, they need to make good on the first stage of an expressivist semantic theory, and tell us what mental states are expressed by complex sentences like “ $\sim P$,” “ $P \& Q$,” and “ s is true,” and to use their answers to those questions, along with background assumptions in the philosophy of mind, in order to predict and explain the commitment tables for “ \sim ,” “ $\&$,” and “true.” So commitment theory doesn’t solve the problems for expressivism—either the Frege-Geach problem or the problem about moral truth. But commitment theory does refine our understanding of the problems facing expressivism, and give us a clearer sense of how, by accomplishing some intermediate goals, an expressivist theory can eventually hope to be able to explain why it is not just a “moral truth” that there is a truth of the matter about whether stealing

money is wrong, and why even moral arguments—and the right moral arguments, at that—can be and are truth-preserving.⁶

It is important to appreciate, however, that in order to do what expressivists need of them, expressivist accounts of truth like the one sketched here in outline cannot just be something to be said about ethics. They have to be understood as quite general theories about the nature of truth—or more precisely, about the meaning of “true.” If the expressivist theory about the meaning of the word “true” does not stand up to the standards requisite for evaluating any other theory of truth, then truth really does create a problem for expressivism, after all.

Fortunately for expressivism, there are at least some good things to be said in favor of the kind of account of truth sketched here. The most general of these is that expressivism is a way of implementing a more general *deflationism* about truth. Recall that expressivists hold that “wrong” does not contribute to what “stealing money is wrong” is *about*, or to the *content* of the thought that stealing money is wrong, but rather contributes to the *kind* of thought the thought that stealing is wrong is. Similarly, an expressivist account of “true” holds that “true” does not contribute to what “s is true” is about, or to the content of the thought that S is true, but rather contributes to the kind of thought the thought that S is true is. This corresponds to the idea that just as wrongness isn’t really a feature of the world, but just a feature of how we think about it, truth isn’t really a feature of the world, either, but just a feature of how we think about it.

Among the many general reasons—applicable well outside of the study of ethics—for thinking that this might be the case about truth, is the famous paradox of the liar. The liar is a sentence which says of itself that it is not true. If truth is a feature of the world, then it would seem that there should be an answer to whether it is a feature that is possessed by the liar, or not. But neither answer seems to be satisfactory—indeed, a seemingly undeniable principle seems to predict that both answers are contradictory. The seemingly undeniable principle which yields this prediction is as follows:

T-schema If S means that P, then S is true just in case P.

The instance of this schema which leads to our problem, is therefore:

T-liar If liar means that liar is not true, then liar is true just in case liar is not true.

Many philosophers have noted not only that the T-schema seems very hard to deny, but two other important things: first, that even if it is not unrestrictedly true, all of its instances exert *pull*, in the sense that they *seem* to us like they have to be true;⁷ and

⁶ See Schroeder (2008b: ch. 11) and especially Schroeder (2010) for a treatment of how expressivists might seek to make good on the first stage of this kind of theory, and explain the commitment tables for “~,” “&,” and “true.”

⁷ See especially Eklund (2002).

second, that we apparently have a *need* for the unrestricted T-schema, since we reason in ways that appear not to make sense if not all instances of the T-schema are true.⁸ This reasoning leads some of those philosophers to believe that every instance of the T-schema *is* true. And that, in turn, leads them to conclude that, since liar does mean that liar is not true, liar is true just in case liar is not true. At worst, following classical logic, this leads some philosophers to conclude that liar is true *and* not true, and at best, some philosophers give up on classical logic, simply in order to stop with a conclusion of the form “P iff \sim P,” rather than getting to an outright conclusion, “P and \sim P.”

Now, it seems like it should be a relatively obvious observation that since even the problematic instances of the T-schema like T-liar seem undeniable, but affirming them leads to paradox, perhaps the right conclusion is that every instance of the T-schema *is* undeniable—though not every instance need be accepted. Perhaps the appropriate attitude toward T-liar is to *reject* it, rather than to deny it. The expressivist account of truth sketched in the last two sections in terms of commitment theory makes good on this idea. It follows from that theory that every instance of the T-schema *is* undeniable. For to deny an instance of the T-schema, one would need to accept the antecedent and deny the consequent. That is, one must accept “S means that P” and deny “S is true if and only if P.” But the commitment table for “true” shows us that someone who accepts “s means that P” has the same commitment for “S is true” as for “P.” Since that is so, she cannot consistently deny “S is true if and only if P.” She can either accept it (if she either accepts or rejects both “P” and “S is true”) or reject it (if she rejects both “P” and “S is true”), but if she denies it, then she is being rationally inconsistent.

So far, this falls far short of a vindication of the expressivist approach to truth; it is more like a reason to wonder whether the expressivist approach might turn out to be defensible on its own merits, after all. And even if the approach has some potential virtues, there is still the first stage of the expressivist theory to be carried out. But promise of being defensible on its own merits is what expressivists *need* of their account of truth. For the noncognitivist project of reclaiming moral truth is not just about moral truth; it is about the whole of truth.⁹

REFERENCES

- Avron, A. (1991). “Natural 3-Valued Logics—Characterization and Proof Theory,” *The Journal of Symbolic Logic* 56: 276–94.
- Ayer, A. J. (1936). *Language, Truth and Logic*. New York: Dover.
- Blackburn, S. (1984). *Spreading the Word*. Oxford: Oxford University Press.
- Blackburn, S. (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.

⁸ See especially Gupta (2005) and Field (2008: ch. 13).

⁹ Special thanks to Michael Glanzburg, Alexis Burgess, Matti Eklund, Barry Lam, Mike McGlone, Jake Ross, and Scott Soames.

- Boghossian, P. (1990). "The Status of Content," *The Philosophical Review* 99: 157–84.
- Dreier, J. (1996). "Expressivist Embeddings and Minimalist Truth," *Philosophical Studies* 83: 29–51.
- Dreier, J. (2006). "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution," *Oxford Studies in Metaethics*, vol. 1. Oxford: Oxford University Press, 217–33.
- Eklund, M. (2002). "Inconsistent Languages," *Philosophy and Phenomenological Research* 64: 251–75.
- Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.
- Geach, P. T. (1958). "Imperative and Deontic Logic," *Analysis* 18: 49–56.
- Geach, P. T. (1960). "Ascriptivism," *The Philosophical Review* 69: 221–5.
- Geach, P. T. (1965). "Assertion," *The Philosophical Review* 74: 449–65.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Gupta, A. (2005). "Do the Paradoxes Pose a Special Problem for Deflationism?" in Jc Beall and B. Armour-Garb (eds.), *Deflationism and Paradox*. Oxford: Oxford University Press, 133–47.
- Hale, B. (1993). "Can There Be a Logic of Attitudes?" in J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection*. New York: Oxford University Press, 337–63.
- Hare, R. M. (1970). "Meaning and Speech Acts," *The Philosophical Review* 79: 3–24.
- Horgan, T. and Timmons, M. (2006). "Cognitivist Expressivism." in T. Horgan and M. Timmons (eds.), *Metaethics after Moore*. Oxford: Oxford University Press, 255–98.
- Horwich, P. (1990). *Truth*. Oxford: Oxford University Press.
- Horwich, P. (1993). "Gibbard's Theory of Norms," *Philosophy and Public Affairs* 22: 67–78.
- Ogden, C. K. and I. A. Richards (1923). *The Meaning of Meaning*. New York: Harcourt Brace.
- Price, H. (1994). "Semantic Deflationism and the Frege Point," in S. L. Tsohatzidis (ed.), *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*. London: Routledge, 132–55.
- Roojen, M. van (1996). "Expressivism and Irrationality," *The Philosophical Review* 105: 311–35.
- Schroeder, M. (2008a). "Expression for Expressivists," *Philosophy and Phenomenological Research* 76: 86–116.
- Schroeder, M. (2008b). *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
- Schroeder, M. (2008c). "What is the Frege-Geach Problem?" *Philosophy Compass* 3: 703–20.
- Schroeder, M. (2009). *Noncognitivism in Ethics*. New York: Routledge.
- Schroeder, M. (2010). "How to Be an Expressivist About Truth," in N. J. Pedersen and C. Wright (eds.), *New Waves in Truth*. New York: Palgrave Macmillan, 282–98.
- Searle, J. (1962). "Meaning and Speech Acts," *The Philosophical Review* 71: 423–32.
- Smith, M. (1994). "Why Expressivists About Value Should Love Minimalism About Truth," *Analysis* 54: 1–12.
- Stevenson, C. L. (1937). "The Emotive Meaning of Ethical Terms." Reprinted in Stevenson (1963), 10–31.
- Stevenson, C. L. (1944). *Ethics and Language*. Oxford: Oxford University Press.
- Stevenson, C. L. (1963). *Facts and Values*. Westport, CT: Greenwood Press.
- Stoljar, D. (1993). "Emotivism and Truth Conditions," *Philosophical Studies* 70: 81–101.
- Wright, C. (1992). *Truth and Objectivity*. Cambridge, MA: Harvard University Press.

CHAPTER 22

TRUTH AND THE SCIENCES¹

ANJAN CHAKRAVARTTY

22.1 INTRODUCTION

WHAT is the relationship of truth to the sciences? Ideally, as a prelude to answering this question, one might begin by specifying what, in this context, truth is generally thought to be. Unfortunately, however, though many philosophers of science believe that the concept of truth is important to a consideration of the practice and outcomes of scientific work, there is no consensus regarding *which* concept of truth, among the several philosophers have developed more generally, is most appropriate here. Indeed, there is no consensus even regarding the question of *whether* the concept is relevant to philosophical considerations of the sciences, and if so, how. In light of this diversity of opinion, perhaps not surprisingly, the task of getting to grips with the relationship of truth to the sciences is not entirely straightforward: it requires an examination of the different approaches to scientific investigation and knowledge that generate these different commitments. In this chapter, I explore the many different ways in which concepts of truth have been brought to bear in connection with these different approaches.

The various attitudes toward truth found in the philosophy of science have correlates in a number of what one might view as philosophically pre-reflective attitudes exemplified in everyday discussions and debates about science. The different philosophical approaches toward the sciences I have alluded to above may thus be thought of as more careful or rigorous philosophical extensions of these everyday attitudes. By way of introduction, then, it may be helpful to enumerate these pre-reflective attitudes before digging into their philosophical counterparts. On the one hand, there are more or less “optimistic” attitudes regarding the connection between truth and the sciences—those that view the sciences as producing truths concerning the natural and social worlds in

¹ For helpful comments on an earlier version of this essay, I am grateful to Agnes Bolinska and Michael Glanzberg.

which we live, or as arbiters with respect to such truths, or at the very least, our best bets for truth production and arbitration. On the other hand, there are more or less “pessimistic” attitudes, which regard the connection as relatively weak or in some cases, illusory. Let us consider this range of possibilities, briefly, in turn.

Almost everyone, I suspect, would agree that the sciences are in the knowledge production business—scientific work produces theories, models, predictions, and explanations which are offered, *prima facie*, as candidates for knowledge. The optimism and pessimism I have gestured toward concern the extent to which these candidates can or should be thought of in terms of truth. Optimists commonly hold that our best scientific theories and models can be thought of as yielding truths regarding their subject-matters, or as furnishing descriptions that are, over time, approaching the truth (in a sense to be specified, as we shall see in sections 22.3 and 22.5). This is the implicit attitude exemplified in everyday contexts in which scientists are presented as authorities regarding everything from the common cold to genetic manipulation to climate change. This generally positive view of the connection between science and truth has its most pronounced philosophical extension in varieties of “realism” with respect to scientific knowledge. Realists of different sorts typically defend the proposition that our best scientific theories provide true or approximately true descriptions of the world, but in different ways: some defend realism rather generally, and others only with respect to certain classes of ontological or other claims which they assess as having greater epistemic warrant than others.

One need not be a realist, however, in order to adopt a positive attitude toward the connection between science and truth. A certain degree of pessimism with respect to truth is compatible with a degree of optimism. There are those, for example, who are somewhat suspicious of the often definitive- and precise-sounding claims of scientists—particularly in connection with things very far removed from our abilities to see or otherwise register information in the absence of elaborate instrumentation. Examples of such things may include the very small, such as viruses and subatomic particles, or the very spatiotemporally distant, such as events in the immediate aftermath of the Big Bang or those leading to the extinction of the dinosaurs. Some skeptics nevertheless hold that scientists are likely on to something, even if they are suspicious about their furthest reaching claims. This everyday attitude also has well-developed philosophical extensions. A number of philosophical positions, including certain varieties of empiricism, regard the sciences as producing truths, but of a restricted subset of scientific claims comprising descriptions of observable phenomena (as opposed to other, less directly detectable phenomena).

Of course, it is possible to be even more strongly skeptical. One might doubt that the sciences produce truths at all, or that while they might do so on occasion, there is little reason to think that they are particularly successful in this regard. Everyday skepticism about the veracity of scientific knowledge has its source in a multiplicity of motivations, from the epistemological to the religious, social, and political. In the philosophy of science, the motivations for this stronger skeptical attitude are invariably epistemological, but in some cases these views do have important social and political dimensions.

Some who ascribe to strongly historicist conceptions of knowledge doubt that truth is especially relevant to an assessment or understanding of scientific knowledge. Social constructivist and feminist critiques of the notion of truth and associated concepts, such as objectivity, engage head-on with the socio-political implications of these ideas, thereby injecting a consideration of them into the epistemology of science.

In what follows, I will outline the range of views that have emerged in the philosophy of science regarding the notion of truth. I will begin in section 22.2 with views that ascribe no role, or a limited role, to truth in connection with scientific practice and knowledge. The primary driver of skepticism about the role of truth here is most commonly manifested as a dissatisfaction with the correspondence theory of truth, according to which truth is a matter of corresponding, in the right sort of way, to a mind-independent reality. (The correspondence theory holds that a belief (statement, proposition, etc.) is true if it corresponds to the way this reality is, or to the facts or states of affairs that make it up; for a detailed treatment of this view, see David (ch. 9 in this volume).) In some of these cases, an account of truth in terms of coherence—i.e. in terms of the consistency of a set of beliefs (statements, propositions, etc.)—may better accord with the relevant approaches to science. (The coherence theory holds that a belief is true if it is part of a coherent system of beliefs; on this view, see Walker (ch. 8 in this volume).) In section 22.3, I will examine positions that view truth in some form as an appropriate aim of science, and in section 22.4, I will consider the even stronger commitment to truth adopted by positions that regard it not merely as an aim, but as an achievement of scientific investigation. In the latter case, truth is generally conceived in terms of some sort of correspondence, or truth-making, or both.

One may think of considerations of truth in connection with the sciences by means of two analytical dimensions: the first concerning the degree to which one may think the notion is applicable to the sciences; and the second concerning the precise theory of truth one may think applies. I will consider, in turn, views that are increasingly interested in truth and, for each, identify the theory or theories of truth that are most appropriate. In section 22.5, I will conclude with an overview of some challenges regarding truth to emerge in recent work, particularly in connection with scientific realism, the most truth-indulgent philosophy of science.

22.2 TRUTH AS ORTHOGONAL TO SCIENTIFIC KNOWLEDGE

22.2.1 Historicism and practice-oriented philosophy of science

Logical positivism (or logical empiricism), the founding movement of the philosophy of science as a recognizably distinct sub-discipline of philosophy, dominated the field

for much of the twentieth century. In the 1960s, the demise of positivism was hastened by a historical turn in the philosophy of science, associated with authors such as Thomas Kuhn, Paul Feyerabend, and Norwood Russell Hanson, which set the stage for many of the issues discussed in the philosophy of science to this day. In particular, Kuhn's extraordinarily influential book, *The Structure of Scientific Revolutions*, played a large role in establishing a form of historicism about scientific knowledge, and an allied (though strictly separable) preoccupation with the actual practice of science by scientists, both of which have significant consequences for thinking about truth. I will take Kuhn's views as exemplary in this regard for present purposes, though it is important to note that in a number of respects, the morals I will draw concerning truth apply not only to some of his contemporaries, but also, in partially overlapping ways, to some of his positivist ancestors.

The underlying commitment of the approach to science taken in the historical turn was to treat the history of the sciences and their constituent practices seriously as a means by which to investigate the nature of scientific knowledge. This was a self-conscious attempt at *descriptive* philosophy—describing the nature of scientific knowledge as it is actually found—as opposed to a normative or more abstract exercise in idealized epistemological speculation. Kuhn presented the history of the sciences as displaying a recurring pattern of development: periods of so-called normal science punctuated by scientific revolutions, which overturned the established order and thereby laid the foundations for a new period of normal science. One might regard classical physics, for example, as spanning a period of normal science, which was ultimately replaced by relativistic and quantum physics, which in part constitute the normal science of our own time. The key implications for truth on this picture derive from Kuhn's characterization of knowledge on either side of a revolutionary divide. Two different periods of normal science, he held, are *incommensurable* with one another and exemplify a phenomenon that is now commonly referred to as *world change*. Let us consider these ideas, briefly, in turn.²

Kuhn held that each period of normal science is typified by a shared commitment to a “disciplinary matrix” or “paradigm,” consisting of a number of elements: symbolic generalizations (such as the mathematical formalism of scientific laws); metaphysical beliefs (for instance, the scholastic commitment to the existence of essential natures); values (such as those favoring certain theoretical virtues, like simplicity); and exemplars (standard problem solving algorithms or techniques). Scientific theories falling under the rubric of different paradigms are incommensurable with one another—a metaphor derived from the origins of the term in the Greek geometrical concept of having “no common measure.” If two theories are incommensurable with one another, they are not incomparable per se, but they are not comparable in a way that would allow one, for instance, to judge that one is true and the other false, or to determine that one is closer to the

² For just a few of the many detailed studies of these topics, see Horwich (1993); Hoyningen-Huene (1993); Sankey (1994); Bird (2000).

truth, relatively speaking. This incommensurability may take the form of a difference in methods or standards operative in science, differences in perception on the part of scientists whose observations are differently “theory laden,” and most importantly for Kuhn, a difference in the very meanings of our words. This semantic incommensurability, which he later analyzed in terms of the failure of meaning-preserving translation (Kuhn 1983), problematizes assessments of truth across paradigms.

Kuhn’s theory of meaning is a species of meaning holism, or meaning contextualism: it suggests that concepts are learned in groups, such as those constituting the sets of interconnected beliefs composing paradigms. As a result, when some of these concepts change, the meanings of terms throughout the network are invariably altered. Thus, the term “mass” as used in classical physics simply does not have the same meaning as the term “mass” as used in relativistic physics, and to suggest that one theory’s characterization of mass is closer to the truth is importantly confused—it is to equivocate with respect to two very different concepts which cannot be understood except from within the paradigms in which they occur. Indeed, this holism or contextualism is so pronounced that Kuhn maintained that there is strong sense in which, after a scientific revolution, scientists live in a different world. In the latter stages of his career Kuhn gave a neo-Kantian gloss to this notoriously cryptic remark. One may think of different paradigms as different lenses one must employ so as to engage in and, indeed, create the reality of scientific phenomena. Perhaps not surprisingly, this is a challenging idea to articulate and defend.

The pursuit of truth plays an important role, on this picture, during normal science. *Within* the scope of a given paradigm, scientists solve problems and generate facts that may be described in terms of truth. But *across* paradigms, such talk becomes incoherent. One cannot judge one scientific theory to be closer to the truth than its predecessor if the very comparison is rendered ineffectual by semantic incommensurability. And if one takes the idea of world change seriously, the problem is compounded, for there is no paradigm-transcendent, God’s-eye point of view from which to judge. There are only paradigmatic judgments; empirical reality is in part structured by our paradigms. One might adopt a coherence theory of truth with respect to the intra-paradigmatic context, since knowledge generated within a paradigm should cohere with it. This will not suffice to allow for anything like scientific progress across paradigms, however, in the way one might hope given a correspondence theory of truth. It is for this reason that those moved by historicist approaches to scientific knowledge are not generally interested in the notion of truth. It is of limited applicability, and entirely inapplicable in the case of correspondence truth. Many who have been inspired by this approach to engage primarily with descriptive features of scientific practice have also found truth to be an unhelpful or irrelevant concept.

22.2.2 The sociology of scientific knowledge

Among other innovations, the historical turn in the philosophy of science turned a spotlight on the social context of the sciences. Descriptions of actual scientific practice

revealed that complex social interactions are intimately connected to the generation of scientific knowledge—from the training of students by mentors, to the collaborative and competitive dynamics of the hierarchies of public and private institutions in which scientific work is done, to the role of economic, political, and other factors driving research by funding and other means, and so on. It was perhaps inevitable, then, that in the wake of the historical turn some should turn to sociology as a means of understanding the sciences. The sociology of scientific knowledge (SSK) investigates the social contexts of the sciences, and in this in itself, it is philosophically neutral with respect to truth. Just as those who are primarily interested in descriptive features of scientific practice might find the notion of truth surplus to analytic requirements, those primarily interested in social aspects might do likewise. In practice, however, most accounts of science inspired by SSK have significant philosophical consequences. Many suggest that once one appreciates the role that social factors play in generating scientific knowledge, a philosophical commitment to some form of *social constructivism* is inevitable, and this once again problematizes the notion of correspondence truth.

The term “social construction” is used somewhat differently by different authors, but let us proceed here with a general understanding of it, according to which it applies to any process wherein what counts as scientific fact (i.e. a claim having the status of scientific knowledge) has been shaped and determined in substantive ways by social factors, and in which different social factors (likely) would have produced facts that are inconsistent with those that were in fact produced. The substantive influences of the social on the scientific are of course numerous (consider, e.g., the directions and methodologies of research permitted, encouraged, and funded), but this does not in itself establish the counterfactual dimension of the idea of social construction—the notion that had such influences been otherwise, one would have ended up with facts that are inconsistent with those currently accepted. In order to motivate the counterfactual, scholars in this vein have typically engaged in case studies of scientific work which aim to show how contingent decisions in the workplace are both determined by social factors and could have gone otherwise, resulting in different and potentially conflicting facts. Additionally, some have offered more global reasons for thinking that such contingency is inevitable. Chief among the proponents of the latter approach are supporters of the so-called “Strong Program” in SSK (also known as the “Edinburgh School”).³

Two of the central tenets of the Strong Program are particularly interesting in connection with the notion of truth. The first is the idea of *symmetry*, which suggests that in the scientific domain inter alia, beliefs that are taken to be true are on a par with those taken to be false in that both have the same causes, and thus should be given symmetrical explanations. This stands at odds with the not uncommon asymmetrical practice of explaining the generation of true beliefs in terms of scientific investigation

³ For a mature statement of the Strong Program, see Barnes, Bloor, and Henry (1996). For a sampling of different approaches to social constructivism, see Knorr-Cetina (1981); Pickering (1984); Shapin and Schaffer (1985); Latour and Woolgar (1986); Collins and Pinch (1993).

successfully latching on to some facet of a mind-independent world, while explaining false beliefs in terms of something going wrong, whether it be instrument malfunction, human error, corruption, or what have you. Putting *all* scientific beliefs on a par with respect to their causes opens the door to a substantively social account of both truth and falsehood. A second tenet of interest is the idea of *reflexivity*, which suggests that the claims of SSK are no less subject to sociological analysis (including the symmetry principle) than any putatively factual claims. This entails a wide-ranging commitment to relativism: truth and falsehood (even with respect to SSK) are defined only in the context of a social community, and have no community-transcendent meaning. As with many relativist positions, there is ongoing debate as to whether this feature of the view renders it self-undermining.

Later incarnations of the Strong Program adopt a theory of language acquisition, *meaning finitism*, which also has profound implications for truth. Finitism can be viewed as an elaboration of the later work of Kuhn, and especially as deriving from the philosophy of the later Wittgenstein. Kuhn emphasized the idea that one learns the meanings of concepts by mastering the use of exemplars that incorporate them, and this point can be extended to training, education, and acculturation more broadly. The ensuing claim of finitism is that meanings are social institutions—the meaning of a term is simply the concatenation of ways in which it can be used successfully in communication within a linguistic community. This relegation of meaning to social facts can be exploited so as to derive a forceful case for the idea that such terms need have no fixed or determinate meanings at all. For on this view, the meaning of a term is constituted by the social acceptance of its use, and this is something that can change subject to social negotiation, reflecting collective decisions about how to go on using language, which may themselves not be universally accepted. This once again appears to favor coherence as an analysis of truth regarding propositions entertained within a scientific community. Whether finitist observations regarding language acquisition successfully entail the associated view of meaning, however, is certainly contestable.

22.2.3 Feminist critiques of science

Just as the historical turn in the philosophy of science suggested and facilitated the emergence of SSK and the development of a number of forms of social constructivism, these latter positions facilitated the development of feminist critiques of science, often in conversation with them. There is significant overlap between SSK and feminist approaches in the acknowledgment of the role of social factors as determinants of scientific fact, but feminist scholars have extended this analysis in more specific ways, reflecting their more specific motivations. Another striking difference is the near universality with which feminist approaches are normative, offering corrective prescriptions for understanding concepts such as objectivity and knowledge (not to mention proposals for reforming scientific practice itself) which, as we shall see directly below, have implications for how one conceives of truth.

Several prominent positions have emerged in this literature, and it will be useful to mention them in passing. *Feminist empiricism* emphasizes the possibility of warranted scientific belief in a communal setting, but crucially, only where biases which enter into research—stemming from gender, ethnicity, socio-economic or political status, and so on—are transparent and appropriately considered. *Standpoint theory* explores the contention that differences in gender, ethnicity, socio-economic and political status, etc. define perspectives to which knowledge is inextricably indexed. *Feminist postmodernism* rejects traditional conceptions of universal or absolute objectivity or truth.⁴ Taken together, these positions offer a number of considerations that might push one in the direction of one theory of truth or another, or indeed, toward a rejection of any substantive theory at all. Foremost among these are considerations of objectivity and perspective. Let us examine them together, as they are interestingly connected.

The term “objective,” when used traditionally in connection with knowledge, has a number of stereotypical connotations. Most important in this context are connotations of disinterest (detachment, lack of bias), and of universality (independence of any particular angle of approach or perspective). Knowledge of this sort arguably comprises propositions that are candidates for describing a mind-independent world, and thus, is consistent with a correspondence theory of truth. There is near consensus in feminist critiques of science that objectivity in the sense of disinterest is regularly violated in science. Case studies are adduced to demonstrate how the presence of, for instance, androcentric bias exemplified by a community of scientists can lead to the establishment of one scientific fact at the expense of another, mutually inconsistent possibility (see, e.g., Longino 1990 for detailed accounts of two cases: one concerning explanations of the evolution of modern human biological characteristics in paleoanthropology; and another concerning the influence of hormones on biological characteristics and behavior in neuroendocrinology). Many also reject the idea of objectivity in the sense of universality or perspective-independence, and as I will suggest, it is this latter thesis that is required in order to problematize the notion of correspondence truth.

Why is the presence of bias in scientific work not in itself sufficient to rule out correspondence? There are two reasons. One is that the presence of bias may not be, in a given case, epistemically significant. That is to say, if regardless of one’s bias one would have assessed the evidence for a given hypothesis in the same way, one might thereby end up with true beliefs concerning that hypothesis in the correspondence sense quite independently of one’s bias. Of course, most feminist case studies are not of this sort—they are cases in which bias *does* make a difference. A second reason that the presence of bias (taken by itself) is consistent with the idea of correspondence truth is that if the institutional infrastructure of science were arranged in such a way as to make

⁴ For a sample of a number of influential versions of these views, see Keller (1985); Harding (1986); Haraway (1988); Longino (1990); Alcoff and Potter (1993); Nelson and Nelson (1996).

epistemically significant bias transparent, then it should be possible in principle to expose such bias and correct for it where necessary. It is for this reason that many feminist philosophers of science advocate for institutional structures that would expose sources of bias. Consider, for example, the idea of effective peer review, where “peers” are drawn from across the space of possible biases exemplified by human agents. The contention that in many and perhaps most domains, the sciences do not yet instantiate such structures, is one motivation for the normative character of most feminist critiques.

If epistemically significant bias is a fact of the matter about much contemporary science, then the prospect of scientific knowledge of a mind-independent world is thereby diminished in practice if not in principle. A more in-principle concern emerges from the idea of perspectives. For just as the Strong Program in SSK embraces a thoroughgoing relativism, where truth is defined only within the confines of a context and not universally, some standpoint theory and certainly all postmodernism is likewise dismissive of the idea of standpoint- or perspective-independent truth. It is important to note here, however, that some would reject this implication. Some standpoint theorists, reflecting a Marxist inspiration, claim that certain perspectives are epistemically privileged over others—typically, subjugated perspectives are privileged over dominant ones as having deeper insight into the subject-matter, in just the way that the proletariat might have deeper knowledge of human potential than the superficial knowledge possessed by those in positions of power. In the absence of such epistemic privilege, however, one is left with an irreducibly fractured picture of warranted scientific belief, where (at best) coherence within perspectives, not correspondence of beliefs from across perspectives, might serve as an analysis of truth.

22.3 TRUTH AS AN AIM OF SCIENCE

22.3.1 The ideal end of inquiry

Having canvassed some reasons for dismissing the relevance of correspondence conceptions of truth in the context of the sciences (in some cases, in favor of a notion of coherence), or for dispensing with concerns about truth altogether, let us turn now to views in which truth is offered explicitly as a goal of scientific investigation. One of the most influential themes to emerge in the philosophy of science in this regard represents scientific inquiry as converging, ultimately, on the truth. The idea of convergence has two features of particular relevance here. One is that convergence is something that can be realized by a process—in this case, the development of scientific knowledge—quite independently of whether the goal itself is ever realized. (Consider the analogy of a mathematical limit that can be approached but never quite realized, except at infinity.) Another feature of interest here is the broad compatibility of the notion of convergence with different views of science. Some scientific realists, for example, whom we will consider

in section 22.4, hold that scientific knowledge is converging on truths in the correspondence sense. But one need not be a realist of this or any other sort to think that truth is the ultimate aim of science.

An important case in point of non-realist endorsement of convergence is associated with the pragmatist tradition in philosophy. The terminology can easily mislead here, since some self-avowed pragmatists also refer to themselves as realists, but as we shall see, their realism is generally not what goes by the name “scientific realism” more specifically. One way of generating the distinction, conveniently for present purposes, is to pay attention to the theories of truth typically endorsed by these camps. While scientific realists generally opt for some version of the correspondence theory—or if they dislike classical accounts of truth, some version of truth-maker theory—pragmatists generally do not. (This is a traditional characterization of pragmatism; for a more nuanced treatment, see Misak, ch. 11 in this volume.) In order to appreciate the pragmatist’s conception of convergence, let us digress momentarily to consider her connected notions of meaning, and truth.⁵

As is the case in any grand philosophical movement, not all participants share precisely the same view, but one tenet shared by many if not all pragmatists is a commitment to a criterion of meaning whereby the content of a proposition is given, as C. S. Peirce put it in his essay “How to Make Our Ideas Clear,” by clarifying its “practical consequences.” Such consequences are to be understood in terms of human experience; that is, as having implications for actual or possible empirical observations. In the context of the sciences, this concerns everything from successful prediction, retrodiction, and communication, to heuristic use and problem solving, where theoretical claims are employed as a basis for action. In the work of William James in particular, this analysis of meaning was parlayed into a theory of truth, according to which positive utility of the sort indicated by useful practical consequences is the marker of truth. Given the fallible nature of scientific knowledge, however, the application of this concept is conceived of by many pragmatists as something of an ideal: truth is whatever it is that will be agreed upon in scientific investigation in the ideal limit of inquiry. For both Peirce and James, the truth in this sense exhausts our conception of reality, which is made by us, not mind-independent. Peirce in particular held that through a process analogous to evolution, scientific methods will converge on a particular body of knowledge. In this way, truth may be conceived as the ultimate aim of science.

The evolutionary analogy here is an evocative one, but its appropriateness in the context of scientific knowledge may be questioned in a number of ways, including some we have already considered. If strongly historicist readings of scientific knowledge are compelling, for example, then the very idea of convergence is problematic, since in this case it is difficult to understand what it could mean for scientific theories across different periods of history to be moving in any particular direction. Indeed, Kuhn himself was

⁵ Here I will focus on some general pragmatist principles drawn primarily from James (1979 [1907]) and Peirce (1998/1992), though readily recognizable in the work of many who have identified with pragmatism since.

attracted to an evolutionary picture of the development of scientific knowledge, but of a rather different sort. On our modern understanding of the analogue—evolution by natural selection—there is no sense in which the adaptations of organisms are “moving” in the direction of anything resembling a definitive ideal. Adaptations in the biological case are more or less optimal only relative to particular conditions of existence represented by particular ecological settings, and the same sort of thing (*mutatis mutandis*) might be thought true of knowledge in historical settings. It is for this reason that many regard the evolutionary analogy as properly suggesting only a notion of progress *from* a particular state of knowledge, as opposed to progress *toward* an ideal, and without the latter, there can be no convergence. Similar appraisals can be generated from the point of view of some social constructivist and feminist positions, and these are contested, as we shall see, by some varieties of realism, which like pragmatism subscribe to the idea of convergence.

22.3.2 Constructive empiricism and one formulation of realism

The pragmatist notion of truth in the ideal limit of inquiry is an example of one manner of thinking about truth as an aim of science. I have described this manner above in terms of conceptions of the *ultimate* aim of scientific investigation. One might think of truth as a goal, however, in rather different terms, as a *proximate* aim. That is to say, one might think that the sciences aim at the truth in the present, not merely as an imagined outcome in a limit. Indeed, it is possible to hold that truth is both a proximate and an ultimate aim of science in precisely these senses—that by aiming at truth in the present, one might thereby realize it in the limit of inquiry, whether this limit is held to be an idealization (as suggested by Peirce), or something one might think is achievable in a finite time, or even now with respect to some scientific facts, as many (if not all) realists contend. This combination of proximate and ultimate aims would seem to be consistent with any position that takes scientific knowledge to be converging on the truth, in some way, shape, or form. Having considered one version of the idea that science aims at truth in the limit, let us turn to the idea that it aims at truth in the here and now.

Bas van Fraassen (1980) has been influential in suggesting that different views of the nature of scientific knowledge can be described in terms of how they view the proximate aim of science. This, I submit, is a tricky business, for it is unclear why *any* view of scientific knowledge should take it to have any one particular or primary aim, given the many candidates that seem to be transparently evident in scientific practice. Based on the activities of scientists, one might plausibly contend that the sciences aim to describe, to predict, to explain, and much more besides. Given this plethora of aims, a question arises as to what philosophical argument could establish the primacy of any one in particular. One possibility here would be to argue that some particular, over-arching,

epistemic aim makes good sense of scientific practice and its various other aims, by accounting for the latter aims as subsidiary to the former. Pronouncements regarding “the (proximate) aim of science” issuing from philosophical accounts of scientific knowledge generally do not engage in such argument, however, and for this reason, any claim of this sort must be viewed as harboring a significant promissory note.

Van Fraassen offers a characterization of scientific knowledge in terms of aims on behalf of two positions: his own view, which he calls “constructive empiricism”; and the view with which he contrasts it, scientific realism. The constructive empiricist holds that the aim of science is *empirical adequacy*, where this is defined in the following way: “a theory is empirically adequate exactly if what it says about the observable things and events in the world, is true” (1980: 12).⁶ In contrast, realism is described as the view according to which the aim of science is truth *simpliciter*—not merely regarding scientific claims about observable things and events, but also regarding claims about unobservable things and events. In both cases, truth may be understood in a correspondence sense, and the distinction between the observable and the unobservable is drawn in terms of human sensory capabilities. Whatever is detectable by the unaided senses (such as planets, alpacas, and metals) is observable, and whatever is not so detectable (such as dark matter, genes, and neutrons) is unobservable. While this particular opposition regarding the putative aim of science has attracted significant attention, one could no doubt imagine other possible epistemological stances with respect to scientific knowledge, each defined by a different proposal for the principal aim of scientific investigation.

There is arguably good reason, however, to be dissatisfied with this strategy for defining epistemological positions such as constructive empiricism and scientific realism. To define an epistemological position purely aspirationally—i.e. in terms of aims—may strike one as too weak. For so defined, these positions are consistent with never actually achieving the relevant aim; indeed, they are consistent with the *impossibility* of achieving the relevant aim. Defined aspirationally, constructive empiricism is consistent with the belief that no scientific theory is empirically adequate, and with the belief that no theory will ever or could ever be empirically adequate. The same moral would seem to apply to realism, in connection with a more comprehensive truth of theories, covering both the observable and the unobservable. This separation of views concerning what science aims to do from those concerning what it actually achieves, or is capable of achieving, is perfectly consistent insofar as one is interested in aims alone. But from the point of view of epistemology, it is too weak. Many realists, for example, believe not only that the sciences

⁶ This definition suffices for present purposes, but it is worth noting that van Fraassen goes on to define empirical adequacy again in more technical terms: a theory is empirically adequate if the observable content of scientific investigation—“the structures which can be described in experimental and measurement reports” (1980: 64)—are isomorphic to empirical substructures of some model or models of the theory, where a model is understood to be any structure of which the axioms and theorems associated with the theory are true. The relevant empirical substructures can then be taken to represent observable things and events.

aspire to truth, but that under certain conditions, they achieve this aim, in a manner to be specified. Let us move on now, to consider philosophical approaches to scientific knowledge that explicitly adopt an achievement view of truth.

22.4 TRUTH AS AN ACHIEVEMENT OF SCIENCE

22.4.1 The many forms of realism

There are many analyses of scientific knowledge that are or could be, if interpreted in a favorable manner, consistent with the idea that scientific theories are true, to some extent or other. This last qualification, “to some extent or other,” introduces a complication which I have thus far ignored, but must now address. Those who believe that the sciences yield truths are often very careful about how they describe this commitment to truth. Some scientific claims (e.g., certain claims about what sorts of things exist), they might say, are likely true *simpliciter*, but others (such as descriptions of laws of nature, or theories taken as wholes) are often merely *approximately true*—i.e. close to the truth, in some sense to be explicated. Convergentsists in this camp then commonly hold that scientific theories are becoming increasingly approximately true over time, with the development of science. I will return to the idea of approximate truth in section 22.5, but let it suffice here to say that when I describe views that take truth to be an achievement of the sciences, I am including those which subscribe to the idea of approximate truth as an appropriate surrogate for truth *simpliciter*, as and where appropriate.

A further qualification is in order here, concerning the breadth of positions that might, on a favorable interpretation, qualify as admitting scientific truths. Many so-called anti-realist positions in the philosophy of science are consistent with scientific truth, but only regarding a very specific domain: the observable. *Instrumentalism*, for example, is the view that scientific theories are simply instruments for predicting observable phenomena, or for systematizing observation reports. On a traditional reading this view also holds that claims pertaining to unobservable things, in themselves, have no meaning at all—they are merely instruments, not candidates for truth or falsity. Less traditionally, positions are sometimes described as instrumentalist if they admit epistemic access only to truths about observables; claims about unobservables here may be true or false, but the relevant truth value cannot be known. In either case, it is open to an instrumentalist to believe that the sciences yield truths about observables in the correspondence sense. Others, such as some logical empiricists, might agree, but only insofar as truth is understood in terms of coherence, in accordance with some or other form of neo-Kantianism. Others, such as Kuhn, might extend the remit of the neo-Kantian-coherence-type approach to truth to the unobservable, but only within periods

of normal science. As should be clear, the possible combinations of restriction on the domain of truth, and theory of truth applicable, are many!⁷

The most unequivocal commitment to truth talk in connection with the sciences, however, is exemplified by versions of scientific realism. Most realists suggest, either explicitly or implicitly, that our best scientific theories are true or approximately true in some appropriate sense of correspondence with a mind-independent world.⁸ Understood this way, scientific realism is the view that scientific theories correctly describe both observable and unobservable features of the world. The main consideration offered in support of realism is an old idea, resonant with common sense, to the effect that the very empirical success of science—in prediction and manipulation of the natural world, not to mention technological applications—indicates that its theories are true. In recent times this idea is commonly referred to as the “miracle argument” (or “no-miracles argument”), after Hilary Putnam’s (1975: 73) contention that realism “is the only philosophy that doesn’t make the success of science a miracle.” In other words, it would be miraculous if the sciences were as empirically successful as they are and yet their theories were untrue, thus suggesting the greater credibility of an explanation of success that appeals to truth. As we shall see shortly, however, the miracle argument is highly contested.

As it happens, there are many varieties of scientific realism, some developed with the intention of responding to forms of anti-realist skepticism such as that concerning the miracle argument. The most general form of realism subscribes to the formula I have already presented: a general endorsement of the truth of our best scientific theories. Of the variations on this formula to emerge, two in particular have come to prominence in recent philosophy of science, both of which offer a more specific prescription for realist commitment. *Entity realism* is the view that under conditions in which one has significant causal knowledge of a putative (unobservable) entity, allowing one (*ex hypothesi*) to manipulate it and use it to intervene in other phenomena, one has good reason to think it exists, and consequently, that claims about the existence of such entities are true. The negative implication of this commitment is that the theories that describe these entities need not be true, and this raises a question about whether an anti-realism about theories is ultimately compatible with a realism about the entities they describe. This view also appears to entail that one may continue to believe in the existence of an entity despite radical changes in one’s (theoretical) understanding of what that entity is like. This in particular might seem problematic from a historicist point of view or, indeed,

⁷ For a formative influence on instrumentalism, see Duhem (1954 [1906]). For discussion of the neo-Kantian reading of logical empiricism, see Richardson (1998) and Friedman (1999), and for a reading more compatible with realism, see Psillos (forthcoming). The neo-Kantian theme is developed differently under the label “internal realism” in Putnam (1981). For a consideration of fictionalism, which holds that things in the world are and behave as *if* our best scientific theories are true (thus admitting truths about observables), see Vaihinger (1923 [1911]) and Fine (1993).

⁸ There are exceptions, however. For arguments in favor of separating realism from the correspondence theory of truth, see Ellis (1988); Leeds (2007); Devitt (2010: esp. chs. 2 and 4). I will return to this idea in another form in section 22.5.

from any point of view that takes accurate description to be an important component of successful reference to an entity.

A second more specific version of realism is *structural realism*, which holds that insofar as scientific theories offer true descriptions, they do not tell us about the natures of things like unobservable entities. Instead, they correctly describe the structure of the unobservable realm. There are two main sorts of structural realist positions: those that take the distinction between structure and nature to be an epistemic distinction; and those that take it to be an ontological distinction. The former suggest that knowledge of the entities that participate in the structural relations described by theories (e.g. by mathematical laws) is simply beyond our grasp, but that theories furnish a kind of structural knowledge nonetheless. The latter suggest that structural knowledge is the most a realist can hope for, because there are in fact no entities that stand in those relations, or that if there are such entities, they are in some sense emergent from or dependent on the structures in which they seem to appear. Like entity realism, structural realism faces interesting challenges. One might wonder, for example, about how clear the distinction between structure and nature is, or whether this distinction is tenable in connection with the concrete systems of things investigated by the sciences. The ontological version specifically faces the challenge of making the relevant ideas of emergence and dependence of entities intelligible.⁹

22.4.2 Skepticism about realism

Entity realism and structural realism each face challenges aimed at the coherence of their specific formulations of realist commitment. Additionally, it is important to note that all the reasons we encountered earlier for being suspicious of the notion of correspondence truth, in connection with different views regarding the nature of scientific knowledge, are also ultimately challenges to scientific realism as it is usually conceived. Often, in the background of these worries, are more general skeptical concerns about scientific knowledge that any realism must face if it is to be defensible. Indeed, the advent of both entity realism and structural realism was intended, in part, to offer responses to some of these more general concerns. I will not here consider whether or to what extent they are successful in this regard, but given the importance of these concerns to motivating (sometimes explicitly but often implicitly) a number of the anti-realist positions we encountered earlier, let us consider them briefly. This more general anti-realist skepticism stems from three notable issues: the ubiquitous use in science of a

⁹ For a defence of scientific realism, see Psillos (1999). For a discussion of novel predictions, often invoked in characterizing what it might mean for something to be one of our best theories, see Leplin (1997). Cartwright (1983) and Hacking (1983) are definitional texts for entity realism, as are Worrall (1989), French (1998), and Ladyman and Ross (2007) for structural realism. For a discussion of the relationships between scientific realism, entity realism, and structural realism, see Chakravartty (2007).

form of reasoning called inference to the best explanation; the so-called underdetermination of theories by data; and discontinuities in scientific theories over time.

Scientific inferences are generally inductive. One is rarely if ever in a position to generate scientific conclusions, laws, theories, and so on, by means of arguments that are deductively valid. Instead, one marshals evidence by means of observation and experiment, and reasons on the basis of this and other commitments to conclusions that are, ideally, inductively strong. In the sciences, a particular form of inductive inference is prevalent, *viz.* inference to the best explanation, according to which one infers hypotheses that, if true, would provide the best explanation for whatever it is that one is attempting to explain. Natural though this pattern of inference may be, anti-realists have been quick to point out two potential difficulties with it in connection with truth. First, in order that one infer the truth, one must have the capacity to rank rival hypotheses reliably with respect to their likelihood of being true. But what assurance is there that the sorts of criteria typically employed in the sciences for purposes of ranking—the simplicity of a hypothesis, its consistency with other hypotheses one may like, its heuristic value, and so on—are indicative of truth? And even if one were to grant that the criteria for choice employed in scientific practice are truth-tracking, in order for inference to the best explanation to result in true belief, it is imperative that the true hypothesis be among those one is considering. But this, arguably, is not generally something one can know in advance.¹⁰

Moving on now to our second skeptical concern, the idea of the underdetermination of theory by data, derived from the work of Pierre Duhem, runs this way.¹¹ Scientific hypotheses do not yield predictions all by themselves. In order to derive predictions, they must be conjoined with background theories, other theories, theories about instruments and measurements, and so on. Now, if observation and experiment result in data that do not agree with one's predictions, what part of one's scientific belief set should one adjust accordingly? Naïvely, one might think that the hypothesis under test should now be viewed suspiciously, but Duhem points out that given all of the things one must assume in order to derive the faulty prediction in the first place, it is not so obvious which part of this apparatus deserves suspicion. Different adjustments to one's belief set—different overall conjunctions of hypotheses and theories—will be consistent with the data. Therefore, there is always more than one overall set of beliefs that is consistent with the data. In more contemporary discussions, underdetermination is usually presented slightly differently, in terms of the idea that every scientific theory has empirically equivalent rivals: ones that agree with respect to the observable but differ elsewhere. This then serves as the basis for a skeptical argument regarding the truth of any particular rival chosen for belief. It is precisely this sort of skepticism that underwrites

¹⁰ For a thorough exploration of inference to the best explanation, see Lipton (2004), and for an influential critique of it, see van Fraassen (1989: ch. 6).

¹¹ The idea of underdetermination is commonly traced to Duhem (1954 [1906]: ch. 6), and many see affinities here with the later "confirmational holism" of W. V. O. Quine (1953). For this reason, it is sometimes called the "Duhem-Quine thesis."

the worries about correspondence truth commonly found in SSK and feminist critiques of science.

Finally, let us turn to an argument that is often simply referred to as the “pessimistic induction,” or the “pessimistic meta-induction.” The history of the sciences is replete with discontinuities in the knowledge they ostensibly provide. Over time, older theories recede into the past as newer theories are proposed and accepted, resulting in changes in belief with respect to the entities, properties and relations, and laws described by these theories. Thus, from the point of view of any given moment in time, most past theories must be considered false, strictly speaking. By induction, then, is it not likely that most present-day theories are also false? The ways in which they will, in future, be revealed as false may not yet be clear to us, but the inductive challenge remains. Given a knowledge of the history of theorizing in any particular domain of scientific investigation, realism may thus appear a rather too optimistic position regarding scientific knowledge. This general skeptical worry can be posed in a number of different ways, focusing on the falsity of past theories, unsuccessful reference on the part of their central terms, and in other ways besides.¹²

Worries about inference to the best explanation, the underdetermination of theory by data, and the pessimistic induction have been highly influential in the philosophy of science. It would be irresponsible not to mention, however, that realists of various stripes have responded to these concerns in a number of ways, and the ensuing debates are very much alive. A thorough canvassing of the details of these debates lies far beyond the scope of this chapter—indeed, these details comprise a large proportion of the field itself. I hope, however, that the ways in which the notion of truth bears on the sciences is now substantially clearer than when we began. My aims in this regard being now largely fulfilled, let us turn in the next and final section to a few important considerations involving truth to emerge in areas of very recent interest in the philosophy of science.

22.5 TRUTH AND SCIENTIFIC MODELS

22.5.1 Abstraction and idealization

One of the most important developments in the philosophy of science over the past two decades has been an increasing awareness and study of the role of scientific models in discussions of scientific knowledge. In keeping with a general movement toward closer attention to scientific practice in the wake of the historical turn, a traditional emphasis on knowledge encapsulated by scientific theories has been supplemented (and perhaps, in some authors’ estimations, replaced) by an emphasis on the ways in which

¹² Canonical formulations of the pessimistic induction are most often drawn from Laudan (1981). For a related and more recent version of history-induced skepticism, see Stanford (2006).

models associated with these theories are used to represent target systems in the world. One reason for this, no doubt, is that “scientific theory” is a somewhat loosely defined concept—its referents vary significantly depending on the context of its use. One and the same theory may be identified in some cases with a set of equations, and in other cases with various models applied to different classes of phenomena putatively described by those equations, but in different ways. These different adaptations of mathematical descriptions to fit particular sorts of cases—sometimes even in ways that are mutually inconsistent—suggest that in many situations, it is the models themselves that are the primary representational tools of scientific description. Thus, a spotlight is turned onto the relationship between modeling and truth.

The term “scientific model” covers many things here. It includes: abstract entities, such as theoretical models about which scientists may have some shared conception; concrete objects, such as graphs, diagrams, other illustrations, and three-dimensional physical representations, such as scale models; and processes, such as computer simulations. Though the forms of these representations are many, the question of how they can be thought to represent their targets veridically, if at all, has become a central topic of debate, not least because, as it happens, scientific models are often constructed so as to *deviate* from the truth. There are two primary ways in which this occurs, and I will call these practices *abstraction* and *idealization*. Although these are terms of art, used in slightly different ways by a growing number of philosophers, there is something of a consensus regarding the basic distinction. For purposes of illustration, I will sketch a representative version of this distinction below.¹³

“Abstraction” and “idealization” can be thought of as labeling processes—ones by means of which models are constructed—as well as the resultant models themselves. Thus, an abstract model is the result of a process of abstraction, a process in which only some of the potentially many factors that are relevant to the nature or behavior of something in the world are represented in a model of it. In the process of abstraction, other factors are ignored, often intentionally, in order to allow for the construction of models that are mathematically or otherwise tractable; and often relevant factors are unintentionally ignored, because their relevance is unknown. Consider, for example, the model of a simple pendulum. Many simplifying assumptions are made in constructing this model, such as the omission of frictional resistance due to the air in which the pendulum swings. Such assumptions are commonly held to compromise the truth of models not only because theoretically important aspects of their target systems in the world are left out, but also because, as a result, the predictions of these models differ from what one finds in reality. One might argue that in both of these ways, abstraction introduces falsehood into practices of scientific representation. By the same token, however, the notion of abstraction suggests one way of making sense of the idea that one representation can be

¹³ For influential discussions in this area, see McMullin (1985); Cartwright (1989: ch. 5); Suppe (1989: 82–3, 94–9). For a comprehensive treatment, see Jones (2005). My take on the distinction here is abridged from Chakravartty (2010a), which considers the implications for truth in some detail.

more approximately true than another: by incorporating a greater number of factors that are causally (or otherwise) relevant to their natures or behaviors.

Similarly, an idealized model is the result of a process of idealization, a process in which at least one of the features of the target system incorporated into the model is represented in a distorted or simplified manner. This process differs from abstraction in that one is not excluding factors, but rather incorporating them in such a manner as to represent them in ways they are not. (In order to sharpen the distinction between abstraction and idealization further, I view the latter in terms of representing things in ways they *could not* be, given the laws of nature. This is sometimes but not always true of abstract models: some abstract descriptions may be true of other systems in the world in which the omitted factors happen not to be present, which is something scientists often contrive in laboratory settings. In any case, given that abstraction and idealization are not mutually exclusive processes, some models may be both abstract and idealized.) Idealized models are, in a stronger sense than in the case of abstractions, false representations of things in the world. It is an implicit assumption of classical physics, for example, that the masses of bodies are concentrated at extensionless points at their centers of gravity, but this is something we clearly know to be false, notwithstanding the predictive and explanatory efficacy of models in classical physics. In such cases, understanding how one model can be more approximately true than another is more complex than in the case of abstraction, turning on how “degrees” of distortion may be conceived in particular cases.

22.5.2 Approximate truth and deflation

The notion of approximate truth is generally viewed as important to scientific realism, versions of which, as we have seen, are commonly more interested to characterize scientific knowledge in terms of truth than any other philosophical assessment of the sciences. The apparent importance of approximate truth here is plain, given that even realists must acknowledge that many if not all scientific theories contain falsehoods and are thus, strictly speaking, false. The task of reconciling the realist’s positive appraisal of scientific knowledge in terms of truth with such obviously recalcitrant evidence of falsity provides strong motivation for thinking about how one theory can be “better off” than another with respect to truth, even when both are false.

The concept of approximate truth, and others such as “verisimilitude” and “truthlikeness” (sometimes used synonymously with “approximate truth” and in other cases not, reflecting the intentions of different authors), have become subjects of detailed and technically striking work in their own right. I will not consider these developments here (but Oddie, ch. 23 in this volume, does). Instead, my goal in this last section is to complete an examination of the implications of the recent interest in scientific modeling for considerations of approximate truth, and for truth more generally. For some advocates of the modeling conception of science hold that once one appreciates how models are used to represent things in the world, all substantive concerns about truth

in this context are dissolved. These concerns arise in the face of two challenges: first, the challenge of explicating the notion of correspondence truth; and second, the challenge of explicating the notion of approximate truth. Given the difficulties inherent in both of these challenges, if it were the case that paying attention to practices of scientific modeling could obviate the need to meet them, this would be a significant result. Let us consider the prospects for doing so.

The most straightforward exposition of the idea that considerations of scientific modeling dissolve substantive issues of truth for the realist is found in the work of Ronald Giere, on whose view a scientific theory is constituted by “two elements: (1) a population of models, and (2) various hypotheses linking those models with systems in the real world” (1988: 85).¹⁴ Hypotheses assert relationships of similarity between aspects of models and their target systems, and are true or false depending on whether these relationships obtain. It is the fact that models are the tools by means of which these epistemic commitments are expressed that dissolves concerns about truth, or so the argument goes. On this approach, Giere contends, there is no question of attempting to forge any sort of correspondence between linguistic entities and the world, and consequently, the challenge to explicate the notion of correspondence truth does not even arise. All that is required are assertions of similarity between models and the world, and for this a deflationary account of truth will suffice.

Deflationism about truth holds that there is nothing metaphysically substantive (such as the notion of correspondence) to the idea of truth (see Azzouni, ch. 17 in this volume). One influential version of the view suggests that the predicate “is true” in “ α is true” is redundant; it adds nothing to the meaning or the assertion of “ α .” Thus, to say that it is true that a two-body Newtonian gravitational model is similar in various ways to the system comprising the earth and moon, for example, is simply to say that these similarities obtain—the truth predicate is redundant. And once scientific claims take the form of asserted model similarities, clarified in terms of respects and degrees of similarity, the need for an overarching account of approximate truth is likewise superfluous, or so one might argue.

The modeling approach facilitates many important insights into the nature of scientific knowledge, but it does not, I suspect, make the issue of correspondence disappear for the realist. Whether one prefers to think of correspondence in terms of a property shared by all truth-bearers, or in terms of the existence of truth-makers for scientific claims, *some* form of correspondence is a requirement of realism. As a quick demonstration of the insufficiency of asserted model similarities in this context, consider the claim that the double-helical structure of Watson and Crick’s physical model of the DNA molecule is similar to the double-helical structure of DNA molecules. Most scientific realists would endorse this claim, but *so too* would most traditional instrumentalists.

¹⁴ Giere refines this portrait of scientific theories in subsequent work, but the refinements are immaterial for present purposes. For his most developed view of scientific knowledge borne of modeling, see Giere (2006), and for an appraisal of this sort of approach, see Chakravartty (2010b). For a longer discussion of modeling, correspondence, and approximate truth, see Chakravartty (2007: sections 7.4–7.5).

The assertion of similarity here does not by itself distinguish the different ontological significance these different parties attach to it. For the realist, the assertion of similarity is interpreted as truly describing the properties of an unobservable entity, whereas for the instrumentalist, it is interpreted as asserting something that is neither true nor false, but useful as a vehicle for the derivation of observable predictions. It is only admitting the relevance of considerations of correspondence, in one form or another, that allows one to distinguish these interpretations.

Neither does the possibility of describing similarities between scientific models and their target systems eliminate the yearning for an account of approximate truth. Though similarities of this sort are the bread and butter of scientific representation, there are undoubtedly other epistemic contexts, of a broader and grander sort, in which scientists and philosophers alike find themselves, understandably, reflecting on the notion of truth. Stepping back from the minutiae of particular representations to the world of scientific knowledge more broadly conceived, there are contexts in which it is entirely appropriate to contend that relativistic physics is closer to the truth than classical physics, or that the modern synthesis in evolutionary biology is closer to the truth regarding the nature of organisms and their populations than biology before Darwin. It is precisely this sense of progress that is celebrated by many realists, and in different ways by other convergentists, and disputed by varieties of historicists and social constructivists. It is this sense of truth that is the subject of so many of our most important and productive philosophical disputes about the relationship of truth to the sciences.

REFERENCES

- Alcoff, L. and E. Potter (eds.) (1993). *Feminist Epistemologies*. London: Routledge.
- Barnes, B., B. Bloor and J. Henry (1996). *Scientific Knowledge*. London: Athlone.
- Bird, A. (2000). *Thomas Kuhn*. Chesham: Acumen.
- Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Clarendon Press.
- Cartwright, N. (1989). *Nature's Capacities and Their Measurement*. Oxford: Clarendon Press.
- Chakravartty, A. (2007). *A Metaphysics for Scientific Realism: Knowing the Unobservable*. Cambridge: Cambridge University Press.
- Chakravartty, A. (2010a). "Truth and Representation in Science: Two Inspirations from Art," in R. Frigg and M. Hunter (eds.), *Beyond Mimesis and Convention: Representation in Art and Science*. Boston Studies in the Philosophy of Science. Dordrecht: Springer, 33–50.
- Chakravartty, A. (2010b). "Perspectivism, Inconsistent Models, and Contrastive Explanation," *Studies in History and Philosophy of Science* 41: 405–12.
- Collins, H. and T. J. Pinch (1993). *The Golem*. Cambridge: Cambridge University Press.
- Devitt, M. (2010). *Putting Metaphysics First: Essays on Metaphysics and Epistemology*. Oxford: Oxford University Press.
- Duhem, P. M. M. (1954) [1906]. *The Aim and Structure of Physical Theory*, trans. P. P. Wiener. Princeton: Princeton University Press.
- Ellis, B. (1988). "Internal Realism," *Synthese* 76: 409–34.
- Fine, A. (1993). "Fictionalism," *Midwest Studies in Philosophy* 18: 1–18.
- Fraassen, B. C. van (1980). *The Scientific Image*. Oxford: Clarendon Press.

- Fraassen, B. C. van (1989). *Laws and Symmetry*. Oxford: Clarendon Press.
- French, S. (1998). "On the Withering Away of Physical Objects," in E. Castellani (ed.), *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*. Princeton: Princeton University Press, 93–113.
- Friedman, M. (1999). *Reconsidering Logical Positivism*. Cambridge: Cambridge University Press.
- Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Giere, R. N. (2006). *Scientific Perspectivism*. Chicago: University of Chicago Press.
- Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.
- Harding, S. (1986). *The Science Question in Feminism*. Ithaca: Cornell University Press.
- Haraway, D. (1988). "Situated Knowledges," *Feminist Studies* 14: 575–600.
- Horwich, P. (ed.) (1993). *World Changes: Thomas Kuhn and the Nature of Science*. Cambridge, MA: MIT Press.
- Hoyningen-Huene, P. (1993). *Reconstructing Scientific Revolutions: The Philosophy of Science of Thomas S. Kuhn*. Chicago: University of Chicago Press.
- James, W. (1979) [1907]. *Pragmatism*. Cambridge, MA: Harvard University Press.
- Jones, M. R. (2005). "Idealization and Abstraction: A Framework," in M. R. Jones and N. Cartwright (eds.), *Idealization XII: Correcting the Model, Poznań Studies in the Philosophy of the Sciences and the Humanities* 86: 173–217.
- Keller, E. F. (1985). *Reflections on Gender and Science*. New Haven: Yale University Press.
- Knorr-Cetina, K. (1981). *The Manufacture of Knowledge*. Oxford: Pergamon.
- Kuhn, T. S. (1970) [1962]. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kuhn, T. S. (1983). "Commensurability, Comparability, Communicability," *PSA* 1982: 669–88.
- Ladyman, J. and D. Ross (2007). *Everything Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Latour, B. and S. Woolgar (1986). *Laboratory Life: The Construction of Scientific Facts*, 2nd edn. Princeton: Princeton University Press.
- Laudan, L. (1981). "A Confutation of Convergent Realism," *Philosophy of Science* 48: 19–48.
- Leeds, S. (2007). "Correspondence Truth and Scientific Realism," *Synthese* 159: 1–21.
- Leplin, J. (1997). *A Novel Defence of Scientific Realism*. Oxford: Oxford University Press.
- Lipton, P. (2004). *Inference to the Best Explanation*, 2nd edn. London: Routledge.
- Longino, H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- McMullin, E. (1985). "Galilean Idealization," *Studies in History and Philosophy of Science* 16: 247–73.
- Nelson, L. H. and J. Nelson (eds.) (1996). *Feminism, Science, and the Philosophy of Science*. Dordrecht: Kluwer.
- Peirce, C. S. (1998/1992). *The Essential Peirce*, ed. N. Houser, C. Kloesel, and the Peirce Edition Project. Bloomington: Indiana University Press.
- Pickering, A. (1984). *Constructing Quarks: A Sociological History of Particle Physics*. Edinburgh: Edinburgh University Press.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.
- Psillos, S. (2011). "Choosing the Realist Framework," *Synthese* 180: 301–16.
- Putnam, H. (1975). *Mathematics, Matter and Method*. Cambridge: Cambridge University Press.
- Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.

- Quine, W. V. O. (1953). "Two Dogmas of Empiricism," in Quine, *From a Logical Point of View*. Cambridge, MA: Harvard University Press, 20–46.
- Richardson, A. W. (1998). *Carnap's Construction of the World*. Cambridge: Cambridge University Press.
- Sankey, H. (1994). *The Incommensurability Thesis*. London: Ashgate.
- Shapin, S. and S. Schaffer (1985). *Leviathan and the Air Pump*. Princeton: Princeton University Press.
- Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press.
- Vaihinger, H. (1923) [1911]. *The Philosophy of "As If"*, trans. C. K. Ogden. London: Kegan Paul.
- Worrall, J. (1989). "Structural Realism: The Best of Both Worlds?" *Dialectica* 43: 99–124. Reprinted in D. Papineau (ed.) (1996). *The Philosophy of Science*. Oxford: Oxford University Press, 139–65.

CHAPTER 23

TRUTH AND TRUTHLIKENESS

GRAHAM ODDIE

TRUTH is widely held to be a cognitive value—perhaps because, other things being equal, it is better to believe a proposition if it is true than if it is false. But even if this is a genuine cognitive norm, it is rather coarse-grained. Not all truths are equally valuable, from a cognitive point of view, and nor are all falsehoods equally disvaluable. The concept of *truthlikeness*, or of *closeness to the truth*, holds out the promise of a richer, more fine-grained classification of propositions, suitable not just for the up or down evaluation of isolated beliefs, but for the calibrated evaluation of cognitive progress in an inquiry.

23.1 COGNITIVE PROGRESS

Consider a simple example. Suppose you are interested in the truth about fundamental particles, and that all you know about such things at the outset of your inquiry are some logical truths, like *either electrons are fundamental or they are not fundamental*. Since these leave the space of possibilities wide open, they do little to help you locate the truth. However, if the standard model is on the right track then learning that *electrons are fundamental* edges you closer to your goal. It is not the whole truth about fundamental particles, but it is a piece of it. If you go on to learn that electrons are one kind of lepton and that all leptons are fundamental, you have edged a little closer. Some truths are closer to the whole truth about fundamental particles than are others.

The discovery that an atom is a composite object displaced the earlier theory that atoms are fundamental. The proposition that protons, neutrons, and electrons are the fundamental components of atoms was then embraced, but it too turned out to be false. Still, this latter falsehood seems closer to the truth than its predecessor, and hopefully the current standard model is closer still, even if it too contains errors, as surely it does. So, some falsehoods may be closer to the truth about fundamental particles than other falsehoods.

Even though a logical truth is not an especially good truth locator, if you moved from *electrons either are or are not fundamental* to the false proposition that *electrons are not fundamental* you would certainly have moved further from your goal. So, some truths are closer to the truth than some falsehoods. But it is by no means obvious that all truths trump all falsehoods. The proposition that electrons, protons, and neutrons are the fundamental components of atoms, though false, may be closer to the truth than the aforementioned logical truth, or even the truth that atoms are not fundamental. If so, certain falsehoods may be closer to the truth than some truths.

These judgments, and many others like them, seem plausible, even platitudinous, but they cry out for a systematic analysis of the central concept—one which would help unify, explain, and vindicate them. This is the logical problem of truthlikeness.

While philosophical attention to the concept of truth goes back millennia, attention to the concept of truthlikeness is a relatively recent phenomenon—beginning in earnest with the pioneering work of Karl Popper in the early sixties,¹ and taking off in several directions when problems with Popper's original account were revealed in the mid-seventies.

That truthlikeness should have made its debut so much later in the game than truth is surprising given the ubiquity of the concept in both everyday and scientific discourse. Perhaps truthlikeness was slow to attract the attention it merits because of a widespread failure to clearly distinguish it from some other truth-related concepts, such as probability and vagueness. But it is demonstrably distinct from both of these.

On almost any account of probability, a logical truth has maximal probability, but it is not at all close to the whole truth. Strengthening a proposition typically lowers probability, but (as we have seen) a sequence of increasingly strong truths may well draw one closer to the truth. And the strongest true theory—the whole truth—will be both the least probable truth and the proposition closest to the whole truth. So in many cases truthlikeness and probability actually pull in opposite directions.

Suppose vagueness is not an epistemic phenomenon, and that it can be explained by treating truth and falsehood as extreme points on a scale of distinct truth values.² Even if this collection of vagueness-induced “degrees of truth” is coherent, they should not be confused with degrees of truthlikeness. Suppose Alan is 179 cm tall. Then the proposition that *Alan is exactly 178.5 cm tall* is not at all vague, so on any degree theory it is *clearly* false. Nevertheless it is close to the truth. That *Alan is tall*, on the other hand, is a vague claim, one that in the circumstances is neither clearly true nor clearly false. However, since it is closer to the clearly true end of the spectrum, it has a high (vagueness-related) degree of truth. Still, *Alan is tall* is not as close to the truth as the quite precise, but nevertheless clearly false proposition that Alan is exactly 178.5 cm tall. So, closeness to the truth and vagueness-related degrees of truth (if there are such) can also pull in different directions.

¹ Popper (1963: 233ff).

² See Smith (2005) for a recent exposition.

Despite its long prehistory of neglect, it is understandable that truthlikeness should have captured Popper's attention, given three of his commitments: that the aim of scientific inquiry is the whole truth of some matter; that most actual scientific theories have either been refuted or are highly likely to be refuted; and that science can, and presumably does, make incremental progress with respect to its central aim of truth. Commitment to this combination of realism, fallibilism, and optimism demands a rich and viable concept of closeness to truth, of truthlikeness.

23.2 POPPER'S ACCOUNT AND THE VALUE OF CONTENT

Popper's essay on the problem was informed by a conviction that logical strength is a cognitive value. His falsificationism deems it a methodological virtue—the stronger a theory is, the easier it is to falsify, and for Popper falsifiability is the hallmark of the scientific. Additionally, however, Popper thought that amongst truths truthlikeness increases monotonically with content. The intuitive examples above seem to bear this out. Call this principle *the value of content for truths* (or *VCT* for short). The problem for Popper was to extend this apparently attractive but limited principle to a more comprehensive ordering, one that embraces falsehoods as well as truths in some non-trivial way.

Popper identified each proposition A with its Tarskian consequence class—the total set of sentences entailed by A in some suitable linguistic framework. The *truth* Popper identified with the complete true theory, T , the set of all true sentences (again, suitably restricted to some linguistic resources that define the matter at issue). The *truth content* of A (A^T) Popper defined as $A \cap T$. *VCT* says both that amongst truths, if $B^T \subseteq A^T$ then A is as close to T as B , and if $B^T \subset A^T$ then A is closer to T than B . Where F is the set of all false sentences A^F (A 's falsity content) Popper defined as $A \cap F$. Popper's proposal is that, *other things being equal*, the larger the truth content, or the smaller the falsity content, the closer a theory is to the truth. That is, A is at least as truthlike as B provided $B^T \subseteq A^T$ and $A^F \subseteq B^F$; and A is more truthlike than B provided, in addition, either $B^T \subset A^T$ or $A^F \subset B^F$.³

Popper's account has some appealing features. In addition to *VCT*, it delivers a non-trivial principle for falsehoods. If A is true and B is false, then A is closer to the truth than B provided A 's truth content entails B 's truth content ($B^T \subseteq A^T$). It follows that if B is false the strongest true consequence B (B^T) is closer to the truth than B itself. This *truth content principle* has been endorsed by every major theory of truthlikeness so far proposed.

Popper's account also has some shortcomings.

³ Popper (1963: 223).

Since a falsehood has false consequences while a truth has none, no falsehood is as close to the truth as any truth. In particular, no falsehood is as close to the truth as a logical truth. As noted, a logical truth leaves the location of the truth wide open, and so should be deemed worthless as an approximation to the whole truth. It follows that on Popper's theory, falsehoods are never more worthwhile than a worthless logical truth. (Dub this the *absolute worthlessness of falsehoods*.)

Further, amongst falsehoods other things cannot be equal: for it is impossible to add a true consequence to a false proposition without thereby adding additional false consequences (or subtract a false consequence without subtracting true consequences). It follows that no falsehood is closer to the truth (or more worthwhile) than any other. This, the Miller-Tichý result, we could dub the *relative worthlessness of falsehoods*.⁴

Finally, it should be *possible* for a truth A to be closer to the truth than its false negation ($\neg A$), even if A falls short of the whole truth. However, unless A is complete ($\neg A$)^T is not a subset of A ^T. So on Popper's account, the only truth closer to the truth than its own negation is the whole truth, T .

Since the first two results involve the falsity content clause, one tempting response is to drop it and measure truthlikeness by truth content alone: A is closer to the truth than B provided $B^T \subset A^T$. Call this the *pure truth content principle*.⁵ This does avoid the absolute worthlessness of falsehoods, but it does so because it implies that the *stronger* of two false theories is the closer to the truth. Call this the *value of content for falsehoods* (VCF). VCF entails, for example, that conjoining any two independent falsehoods is *always* a step closer to the truth than either of the conjuncts. Strengthening a falsehood may well be a step toward the truth, but it is not guaranteed to be such. VCF has seemed highly problematic to nearly every investigator in this domain, the most notable exception being David Miller.⁶

An influential categorization of post-Popper approaches to the problem divides them into *content* and *likeness* camps.⁷ The content approach, which takes its inspiration from Popper's assay, takes truthlikeness to supervene on two factors: a truth factor and a content factor. By contrast, the likeness approach takes truthlikeness to supervene on something that doesn't appear in Popper's account—likeness or distance between possibilities. However, a third approach, one which is also inspired by Popper's account, can and should be distinguished from both of these—namely, the *consequence approach*. There are doubtless other ways of conceptualizing the options: bottom-up versus top-down; quantitative versus qualitative; and syntactic versus semantic, and so on. However, the content-consequence-likeness trichotomy is a convenient place to start.

⁴ Miller (1974); Tichý (1974). Let A and B be false and assume $A^F \subset B^F$. Consider b such that $b \in B^F$ and $b \notin A^F$, and let $a \in A^F$. Clearly $(\neg a \vee b) \in B^T$. If $(\neg a \vee b) \in A^T$ then we have $b \in A^F$: contradiction. So not $B^T \subseteq A^T$.

⁵ Popper (1976).

⁶ For a recent intriguing and feisty defense of VCF, however, see Miller (2007: 185).

⁷ See Zwart (2001).

23.3 THE CONTENT APPROACH

According to the content approach, truthlikeness supervenes on two factors: a *truth* factor, and a *content* factor. Amongst propositions with the same truth factor, truthlikeness varies with content alone; and amongst equally contentful propositions, truthlikeness varies with truth value alone. However, this supervenience thesis is too weak to fully characterize the approach. For example, it would count the following theory within the content camp: *truthlikeness decreases with increasing content*. This account yields an ordering that is not only insensitive to truth but is exactly the reverse of *VCT*—thus conflicting with the guiding desiderata of the content approach. We need an additional constraint on supervenience, and the simplest is clearly *VCT* itself. *VCT* is the core content ordering and every content ordering extends it, while satisfying supervenience on truth and content. Popper's qualitative account clearly falls within the content approach so characterized.⁸

David Miller has proposed a general theory of propositional distance that yields a theory of truthlikeness that can also be placed within the content approach. Although Miller's account is inspired by the general algebraic approach to logic and is not tied to any particular account of propositions, it is heuristically useful to frame it within a possible-worlds framework. Let a proposition be identified with the class of worlds in which it is true, and assume further that the class of all propositions forms a complete Boolean algebra.⁹ The conjunction of two propositions is their intersection, disjunction is union, negation is set-theoretic complementation, and entailment is the subset relation. A proposition is true if it contains the actual world, and the truth, T , is the singleton of the actual world. The set-theoretic *distance* between propositions A and B is defined as their symmetric difference $(A \Delta B) : (A - B) \cup (B - A)$. The subset (entailment) relation \subseteq then serves to partially order these distances: A is as close to C as B if $(A \Delta C) \subseteq (B \Delta C)$, and closer if $(A \Delta C) \subset (B \Delta C)$. Note that we have qualitative analogues of the three familiar properties of a numerical distance function. (i) $A \Delta B = \emptyset$ if and only if $A = B$; (ii) $A \Delta B = B \Delta A$; (iii) $(A \Delta B) \cup (B \Delta C) \subseteq (A \Delta C)$. The distance of A from the truth T is $A \Delta T$. A is as close to the truth as B if $(A \Delta T) \subseteq (B \Delta T)$, and closer if $(A \Delta T) \subset (B \Delta T)$. If A is true then $A \Delta T$ is $A - T$; if false, $A \cup T$. So whether A and B are both true or both false: $(A \Delta T) \subseteq (B \Delta T)$ just in case $A \subseteq B$ (and $(A \Delta T) \subset (B \Delta T)$ just in case $A \subset B$.) The account thus delivers *VCT* as desired, but also delivers *VCF* along with it. Further, like Popper's theory, it deems no incomplete truth closer to the truth than its negation.

This qualitative ordering can be extended to a quantitative one, by means of a countably additive, strictly positive measure μ on the logical space. If the distance between A and

⁸ In Oddie (2013) a slightly more complicated characterization is employed, but it yields the result that *VCT* is the core content ordering, and that every content ordering extends *VCT*.

⁹ Note that we need not hold that any arbitrary class of worlds defines a proposition, but for completeness the intersection of any class of propositions has to be a proposition.

B , $\delta(A\Delta B)$ is $\mu(A\Delta B)$ then δ possesses the three properties of a metric: (i) $\delta(A\Delta B) = 0$ if and only if $A=B$; (ii) $\delta(A\Delta B) = \delta(B\Delta A)$; (iii) $\delta(A\Delta B) + \delta(B\Delta C) \geq \delta(A\Delta C)$. The distance of A from the truth is $\delta(A\Delta T)$. So:

if A is true, $\delta(A, T) = \mu(A) - \mu(T)$;

if A is false, $\delta(A, T) = \mu(A) + \mu(T)$.

If we take the content of A , $Ct(A)$, to be $(1 - \mu(A))$, and the truthlikeness of A , $Tl(A)$, to be $(1 - \delta(A, T))$ this is equivalent to:

if A is true δ , $Tl(A) = Ct(A) + \mu(T)$;

if A is false, $Tl(A) = Ct(A) - \mu(T)$.

making it explicit that on this proposal truthlikeness is an additive function of a content factor and a truth factor. Further, if A and B have the same truth value, A is more truthlike than B if and only if $Ct(A) > Ct(B)$ —i.e. the quantitative analogue of *VCT* and *VCF*. Miller has extended and refined this proposal within a general algebraic approach to logical distance in a number of papers, and most recently, in Miller (2010), he has extended it well beyond classical Boolean algebras. Further, Miller's approach has been the springboard for interesting developments by Mormann and Gerla, which embed this algebraic approach within the general theory of metric spaces.¹⁰

Theo Kuipers has also developed a symmetric difference proposal formally similar to Miller's, but within a structuralist paradigm.¹¹ Kuipers takes a theory to be a class of structures—all those the theory deems nomically possible—but unlike standard possible-worlds semantics, the true target theory is not a singleton. Rather the target theory, T , is the set of *all* nomically possible structures, given the actual laws, rather than a single structure complete in both nomic and non-nomic details. Kuipers presents two accounts: "naïve" and "refined." The naïve account is formally identical to Miller's symmetric difference proposal, but with the caveat that T is not a singleton. As in standard possible-worlds accounts, a theory is true just in case it is a superset of T and otherwise it is false. *VCT* follows, for if $T \subset A$ (A true) and $A \subset B$ (A stronger than B), then $A\Delta T \subset B\Delta T$ (A closer to T than is B). However, because T is a possibly large class of structures, we get quite a different account of falsehood, and in virtue of this *VCF* is avoided. For example, suppose C and D are two independent, obtaining, non-nomic states, the negations of which are nomically possible. Then $T \wedge C \wedge D$ and $T \wedge C$ both exclude some nomic possibilities, and so neither is "true." Further, $(T \wedge C \wedge D) \subset (T \wedge C)$ (i.e. $T \wedge C \wedge D$ entails $T \wedge C$), while $((T \wedge C)\Delta T) \subset ((T \wedge C \wedge D)\Delta T)$, rendering $T \wedge C \wedge D$ further from T than $T \wedge C$. So, strengthening a "falsehood" on this account does not

¹⁰ See e.g. Miller (1977); Mormann (2005; 2006); Gerla (2007).

¹¹ Kuipers (1987; 1992; 2000).

necessarily increase truthlikeness (though of course it may). If we extend this qualitative ordering to a quantitative account by measuring the size of the symmetric difference, then it is clear that this account doesn't fall squarely within the content approach as Miller's does (among false theories truthlikeness does not vary with degree of content alone). However, Kuiper's naïve symmetric difference account still yields something rather close to *VCF*. Call a theory *A* that rules out *all* the *T*-possibilities *strongly false*. (Note that in the classical case, where *T* is a singleton, strong falsity is just classical falsity.) Then the naïve account yields the *value of content for strong falsehoods*: the logically stronger of two strongly false theories is closer to the truth.

These results might suggest that, given supervenience on truth and content, any extension of the value of content for truths into the realm of false propositions will entail the value of content for falsehoods, or something close to it. But that is too swift, and indeed Popper's account (which also extends *VCT* within the content approach, albeit minimally) is a counterexample. There are three well-behaved ways for the truthlikeness of falsehoods to vary with content: increase with increasing content; remain constant; or decrease with increasing content. If truth and content are separable, or independent, value factors then (given some weak continuity assumptions) truthlikeness will be an additive function of the two factors.¹² If truth and content are additive contributors to truthlikeness, and if truthlikeness increases with content among truths (*VCT*), then truthlikeness will also increase with content among falsehoods. So, additivity together with *VCT* entails *VCF*. Given that *VCF* is problematic, the content theorist has to embrace a non-additive function of truth and content.

Suppose truthlikeness among falsehoods remains constant—i.e. all falsehoods have the same degree of truthlikeness. Then we get the *relative* worthlessness of falsehoods: no falsehood is more truthlike than any other.¹³ Suppose that truthlikeness is a decreasing function of content among falsehoods. The weakest falsehood is the negation of the truth $\neg T$, which would thus be the most truthlike of falsehoods. By the truth content principle (for false *A*, $A \vee T$ is closer to the truth than *A*) the truth content of $\neg T$ would be more truthlike than $\neg T$ itself. The truth content of $\neg T$ is $(\neg T \vee T)$, a logical truth. Given *VCT*, a logical truth is the least truthlike truth. Thus the most truthlike falsehood would be further from the truth than the least truthlike truth. That is, we have the *absolute* worthlessness of falsehoods.

It appears that to avoid the value of content for falsehoods, a content theorist is forced to endorse one or other of two features that hobbled Popper's account: the relative or absolute worthlessness of falsehoods.

¹² For the connections between independence of factors and additivity in the context of measurement theory, see Krantz et al. (1971).

¹³ As noted in section 23.9, Popper's ordering is realized by a non-additive function of truth and content, one that renders all falsehoods equally distant from the truth.

23.4 THE RELEVANT CONSEQUENCE APPROACH

These problems with the content approach have suggested a different development of Popper's central hunch. According to the *relevant consequence* approach, Popper was right that truthlikeness supervenes on true and false consequence classes, but wrong in thinking that all consequences of a theory count. Some are *relevant*, some are not.¹⁴

Examples from the familiar weather framework (which features the three primitive weather states: *hot*, *rainy*, and *windy*) bear this out. Suppose it is hot, rainy, and windy ($h \wedge r \wedge w$). The truth that it is rainy and windy ($r \wedge w$) seems closer to the whole weather truth than the falsehood that it is dry and still ($\neg r \wedge \neg w$). Further, this superiority seems to be preserved under the addition to each of the falsehood that it is cold: ($\neg h \wedge r \wedge w$) seems closer to the truth than ($\neg h \wedge \neg r \wedge \neg w$). Two truths count in favor of ($\neg h \wedge r \wedge w$) (i.e. *rainy*, *windy*) and one falsehood counts against (*cold*). No comparable truths count in favor of ($\neg h \wedge \neg r \wedge \neg w$) while three falsehoods (including *cold*) count against.

There are of course other true consequences of ($\neg h \wedge \neg r \wedge \neg w$)—like ($h \vee \neg w$)—and it is just such consequences that fuel the Miller-Tichý incomparability proof.¹⁵ But a true consequence like ($h \vee \neg w$) does not enhance the cognitive value of ($\neg h \wedge \neg r \wedge \neg w$). This suggests that some consequences are relevant to the determination of truthlikeness while some are not. Other considerations also suggest this. Two false propositions of the same degree of strength (such as ($\neg h \wedge r \wedge w$) and ($\neg h \wedge \neg r \wedge \neg w$)) entail the same number of truths as each other, and the same number of falsehoods as each other. (And every falsehood entails the same number of truths as it does falsehoods.) So letting all consequences count equally would render all equally strong false propositions equidistant from the truth, and trading off false consequences against true consequences would render all falsehoods equally far from the truth (relative worthlessness).

These considerations motivate a *relevance* constraint on truth and falsity content. Let R be a criterion of the relevance of consequences, and let A_R be the set of consequences of A that satisfy R . What counts as a relevant consequence varies from one account to another, but whatever R is, A_R has to be equivalent to A itself. (Popper's account is the limiting case in which R is the vacuous criterion P which all consequences satisfy: $A_P = A$.) A_R^T (the R -relevant truth content of A) is defined as $A_R \cap T$; and A_R^F (the R -relevant falsity content of A) is defined as $A_R \cap F$. Given bivalence, $A_R^T \cup A_R^F = A_R$,

¹⁴ Although this basic idea was first explored in Mortensen (1983). Mortensen abandoned the proposal by the end of the article. Schurz and Weingartner (1987; 2010) have articulated an adequate version of the theory for propositional languages, one which is related to Gemes (2007) more general proposal. Gerla (2007) uses the notion of the relevance of a "test" or factor, but his account is best located more squarely within the likeness approach. Incidentally Gerla also points out (729) that "Popper has acknowledged (in his 1976) that 'the problem of truthlikeness can perhaps be solved only by a relativization to relevant problems.'"

¹⁵ See fn. 4.

so the union of the relevant truth and falsity contents of A is also equivalent to A . If A is true, A_R^F is empty, and $A_R^T = A_R$. Where P is the vacuous criterion, A_P^T is the set of all A 's true consequences, and A_P^F is the set of all A 's false consequences.

Schurz and Weingartner (2010) use the *entailment* relation among relevant consequence classes, rather than the subset relation, to define closeness, but otherwise their qualitative account tracks Popper's: A is as close to the truth as B if $B_R^T \subseteq A_R^T$ and $A_R^F \subseteq B_R^F$, and is closer if in addition either $B_R^T \subset A_R^T$ or $B_R^F \subset A_R^F$. This yields Popper's account as the limiting case where $R = P$.

A possible relevance criterion that the *h-r-w* examples might suggest is *atomicity*. However, since many logically distinct propositions imply the same atomic propositions, atomicity would not secure the equivalence of A and A_R —both $(h \vee r)$ and $(\neg h \vee \neg r)$ for example imply no atomic propositions. Schurz and Weingartner, exploiting properties of the conjunctive normal form theorem for propositional languages, opt for the following relevance condition S : *being picked out by a disjunction of atomic propositions or their negations* (where no disjunct is redundant). This secures a range of the intuitive judgments, such as:

$$(h \wedge r) >_S (h \wedge \neg r) >_S (\neg r) >_S (\neg h \wedge \neg r), \text{ and}$$

$$(h \vee r) >_S (\neg h \vee \neg r) >_S (\neg h \wedge \neg r).$$

Using S -relevance avoids relative worthlessness (e.g. $(h \wedge \neg r) >_S (\neg r)$), and allows an incomplete truth to be closer to the truth than its negation (e.g. $(h \vee r) >_S (\neg h \wedge \neg r)$). Clearly S -relevance avoids VCF (e.g. $(\neg r) >_S (\neg h \wedge \neg r)$). And for every R , the relevant consequence approach yields VCT . For suppose A and B are true and A strictly entails B ($B^T \subset A^T$). Then since A_R is equivalent to A , and $A_R = A_R^T$, A_R^T is equivalent to A , and B_R^T is equivalent to B . So A_R^T strictly entails B_R^T . And since both A_R^F and B_R^F are empty, B_R^F entails A_R^F : $A >_R B$.

A closely related proposal is Burger and Heidema's (1994) comparison of theories by *positive and negative components*. A positive sentence is one that can be constructed out of \wedge , \vee , and any true basic sentence (where a basic sentence is either an atomic sentence or its negation). A negative sentence is one that can be constructed out of \wedge , \vee , and any false basic sentence. Call a sentence *pure* if it is either positive or negative. If we take the relevance criterion to be *purity*, and combine that with the relevant consequence schema, we have Burger and Heidema's proposal, which yields a good number of the intuitive judgments like those listed above. Unfortunately purity (like atomicity) does not quite satisfy the constraint that A be equivalent to the class of relevant consequences. For example, if h and r are both true then $(\neg h \vee r)$ and $(h \vee \neg r)$ both have the same pure consequences (namely, none).¹⁶

The relevant-consequence approach faces three major objections.

¹⁶ Burger and Heidema (1994) consider a refinement which uses convex and non-convex components, but it is not easy to fit this version into the relevant consequence approach.

The first is an *extension problem*: the approach may produce acceptable results in a finite propositional framework but it needs to be extended to more realistic frameworks—e.g. first-order and higher-order frameworks. A recent proposal by Gemes (2007) is promising in this regard. More research is required to demonstrate its adequacy.

The second is the problem of worthlessness: no false proposition is closer to the truth than any truth, including logical truths. Schurz and Weingartner have answered this objection by extending their qualitative account to a quantitative account, by assigning weights to relevant consequences and summing. On their quantitative account $(h \wedge r \wedge \neg w) >_s (h \vee \neg h)$, for example.

The third involves the language-dependence of any adequate relevance criterion. This problem will be outlined and discussed below (section 23.8).

23.5 THE LIKENESS APPROACH

Assume that for the purposes of the logical problem of truthlikeness, a proposition can be identified with the class of worlds with which it is compatible. Without additional structure on the logical space we have two factors for a content theorist to work with—the size of a proposition (content factor), and whether it contains the actual world (truth factor). The likeness approach requires some additional structure to logical space—namely, that worlds are ordered according to their differing *likeness to or distance from* each other. The core of the likeness approach is that the truthlikeness of a proposition supervenes on the likeness or the distance between worlds in logical space.

The likeness theorist has two initial tasks: first, making it plausible that there is an appropriate likeness or distance function on worlds; and second, extending likeness between individual worlds to likeness of propositions (i.e. sets of worlds) to the actual world.

Tichý (1974) proposed a simple account of likeness between worlds in a finite propositional framework. Each world can be associated with a unique collection of atomic states—all the atomic states that are true in that world. The symmetric difference operation on worlds, so construed, yields the collection of states true in one world but not in the other—that is, the class of atomic states over which the two worlds disagree. As with Miller's symmetric difference on propositions, a partial ordering of worlds by distance (or likeness) is induced by the subset relation on symmetric differences: world u is closer to w than is v just in case $(u \Delta w) \subset (v \Delta w)$. If, furthermore, atomic states have weights attached then the distance between two worlds can be defined as the total weight of their symmetric difference. (Tichý assumes that atomic states all have the same weight, but that simplification can be easily dropped in favor of differential weightings.) Tichý also proposes a simple extension of distance between worlds to propositional distances from the actual world: namely the *average* distance from the actual world of the worlds in the proposition. (Tichý assumes that worlds have the same weight, but again that is gratuitous and eliminable. We can substitute a different weighting of worlds, if such be

justified, and take distance from the truth to be *weighted average* distance from the actual world.)

An early objection to this account (made by Popper himself) is that it violates Popper's core desideratum: *VCT*.¹⁷ Suppose we have a framework with n atomic propositions (a_1, \dots, a_n) all of which are true and consider the following rather odd pair of truths:

- A *Either* all atomic propositions are true *or* all are false.
- B *Either* all atomic propositions are true, *or* all are false, *or* all are true except a_1 which is false.

Assuming all atomic propositions have the same weight ($1/n$) the average distance of worlds in A from the truth actual world is $\frac{1}{2}$, whereas the average distance of worlds in B from the actual world is $(n+1)/3n$, which, provided $n > 2$, is smaller than $\frac{1}{2}$. But A and B are both true and A is the logically stronger. So averaging distances, where the world-distances themselves are generated by symmetric difference on atomic propositions, violates *VCT*. Weakening a proposition by adding worlds sufficiently close to the actual world reduces average distance from the truth. Popper naturally took this to be a knock-down refutation, and many others have concurred.

In his (1976), Hilpinen articulated a likeness account independently of Tichý, but without defining or presupposing a measure of likeness/distance between worlds. Instead, utilizing Lewis's apparatus of nested spheres centered on the actual world, he assumes a total likeness ordering of worlds with respect to the actual world. He extends this world-ordering to propositions using two parameters: the largest sphere of worlds that does not overlap with A ($\min(A)$); and the smallest sphere of worlds that contains A ($\max(A)$). Hilpinen located his account within Popper's approach, characterizing $\min(A)$ as a truth factor, and $\max(A)$ as a content factor. It is true that $\min(A)$ is a measure of *closeness to being true*, since the smaller $\min(A)$ is, the closer the boundary of A is to the actual world.¹⁸ If a proposition is true *simpliciter* then $\min(A)$ is empty, and A cannot be any more true. So all truths are on a par as far as this factor goes, but falsehoods clearly differ. $\min(A)$ is thus not simply a truth factor. It is something more than the truth factor. $\max(A)$, on the other hand measures the distance of the outer boundary of A from the actual world, and it does yield a kind of content ordering: if A entails B then $\max(A) \subseteq \max(B)$. Still, it is a very crude content measure, as we will see. $\max(A)$ is something less than the content factor. According to Hilpinen, A is as close to the truth as B if $\min(A) \subseteq \min(B)$ and $\max(A) \subseteq \max(B)$, and closer if either $\min(A) \subset \min(B)$ or $\max(A) \subset \max(B)$.

Consider a complete proposition W (for *Worst*) which contains the world farthest from the actual. (W seems to be the least truthlike proposition on offer.) W is among

¹⁷ Popper (1976).

¹⁸ Weston (1992) essentially defines the related notion of *approximate truth* in terms of the sufficient smallness of $\min(A)$: "The basic idea . . . is that a statement will be approximately true (in) I if there is (a) J which is 'near' I, and (in) which it is actually true' (61–2).

the strongest propositions there are, while $(W \vee \neg W)$ is the weakest. Nevertheless $\max(W) = \max(W \vee \neg W)$. So \max is a crude indicator of content. As a corollary, this account does not deliver the full value of content for truths. Truth A can be logically stronger than truth B while $\min(A) = \min(B)$ and $\max(A) = \max(B)$. For example, $(W \vee T)$ is among the strongest truths, while the tautology, $(W \vee \neg W)$ is the weakest. $(W \vee T)$ and $(W \vee \neg W)$ have the same truth factor and the same content factor, and so are deemed equidistant from the truth. It does, however, yield that if A and B are true and A entails B then A is not further from the truth than B . Note further that while the *min-max* proposal avoids the relative worthlessness of falsehoods it endorses the absolute worthlessness of falsehoods—no false proposition is closer to the truth than any truth—because if A is true and B is false $\min(A) \subset \min(B)$. This defect can be overcome if we replace the distance ordering with a distance measure on worlds, and take the average of the extreme worlds to measure overall distance.¹⁹

Suppose we have a partial, rather than total, likeness ordering on worlds. We cannot extend such an ordering to all propositions by *min-max* since that presupposes a total ordering. Consider a condition which in the special case of a total ordering of worlds is equivalent to the *min-max* extension: $A \geq_m B$ if and only if for every world in B there is a world in A that is at least as close to the actual world, and for every world in A there is a world in B that is at least as far from the actual world. (And let $A >_m B$ be the corresponding condition if one of those inequalities is strict.) Interestingly if we assume the partial ordering of worlds induced by the symmetric difference on atomic states and extend it by this condition then what we end up with is Burger and Heidema's (1994) relevant consequence ordering where purity is the relevance criterion, placing that ordering within the likeness approach also.

In a series of papers, and a magisterial book on the topic, Niiniluoto has urged a different extension of likeness between worlds to truthlikeness of propositions, one which is within the spirit of Hilpinen's, but which overcomes the coarse-grained nature of Hilpinen's content factor. Assume we have numerical distances, δ , between worlds, and let w be the actual world. $\min(A)$ is distance from actuality of worlds in A closest to actuality. Now consider the sum S of the distances of all worlds from the actual world and let the weight $\mu(u)$ of world u be $\delta(u, w)/S$. Finally, let $\text{sum}(A) = \sum_{u \in A} \mu(u) = \sum_{u \in A} \delta(u, w) / \sum_u \delta(u, w)$. $\text{sum}(A)$ is a regular normalized measure, but one which is sensitive to the distances of worlds in A from the actual world. If worlds in A are close to the actual world, $\text{sum}(A)$ is small; if far, then $\text{sum}(A)$ is large; and the logically weaker A is, the larger its sum . Hence. Unlike \max , $\text{sum}(W \vee T)$ is much smaller than $\max(W \vee \neg W)$. Distance from the truth Niiniluoto defines as a variably weighted average of *min* and *sum*, the relative weights serving as indicators of a preference for truth or for content.²⁰ But so long as both truth and

¹⁹ Gerla (2007) arrives at an extension within a somewhat different framework which is close to numerical *min-max*.

²⁰ See Niiniluoto (1987; 1998).

content are given some weight, *min-sum-average* delivers the full value of content for truths.

There are three major objections to the likeness approach, which parallel objections to the relevant consequence approach: an *extension* problem, a *complexity* problem, and a *language dependence* problem.

23.6 THE EXTENSION PROBLEM

Schurz and Weingartner argue that the main problem with the likeness approach is that “the problem of extending truthlikeness from possible worlds to propositions is intuitively underdetermined.”²¹ Even if we are granted an ordering or a measure of distance on worlds, there are different ways of extending that to propositional distance, and no objective way to decide between them.

One way of answering this objection head-on is to identify principles that, given a distance function on worlds, constrain the distances between worlds and sets of worlds, principles perhaps powerful enough to identify a unique extension.

First, suppose that all the worlds in A are the same distance from the actual world. What is the overall distance of A from the actual world? The natural answer is that A is the same distance as the worlds it contains:

The uniform distance principle:

If the worlds in A are of a uniform distance d from the actual world, then the distance of A from the actual world is also d .

Let $A^{v/u}$ be any proposition that differs from A only in that it contains v rather than u , and suppose that v is closer to the actual world than u . Clearly $A^{v/u}$ cannot be *further* from the actual world than A is. (Note that Tichý’s, Hilpinen’s, and Niiniluoto’s likeness measures all deliver this.) That is,

The pareto principle:

If v is at least as close to the actual world as u is, then $A^{v/u}$ is at least as close to the truth as A is.

If v is closer to the actual world than u is then there should be a difference between the distance of A from the truth and the distance of $A^{v/u}$ from the truth. What should that difference depend on? Given that $A^{v/u}$ differs from A only over the distance from the

²¹ Schurz and Weingartner (2010: 423). They go on: “. . . and it seems that this extension problem is the root of the conflict between content- and likeness-considerations as diagnosed by Zwart and Franssen 2007.”

actual world of worlds u and v , the difference in closeness to truth of A and B can certainly depend on the distance of u from the actual world and the distance of v from the actual world. The following argument shows that the difference may also depend on the size of A . The smaller A is the more the replacement changes what we might call A 's *distance profile*. In the limit if A is a singleton (*viz.* $\{u\}$), $A^{v/u}$ is also a singleton (*viz.* $\{v\}$). From the uniform distance principle, we know that the difference between the distances of A and of $A^{v/u}$ in this case is the difference between the distance of v and the distance of u from the actual world. And that is the *largest* difference that replacing u with v could make. The larger A is the less of an impact the replacement will have. So size of A may make a difference to the impact of replacement. However, we don't have to stipulate any particular function here, or even that it be a decreasing function of the size of A (as indeed it should be). Rather, we merely allow that the difference between the two distances is some function or other of these three factors:

The difference principle:

The difference in closeness to the truth of A and $A^{v/u}$ is some function or other of at most three factors: the distance of u from the actual world; the distance of v from the actual world; and the size of A .

These three conditions, which we will call *the distance extension principles*, are individually attractive and rather weak. It is easy to check that *average* satisfies the distance extension principles, but it can also be shown that *average* is the only extension to do so. Any other extension will violate one of the three constraints.²²

As critics have often pointed out, there is something rather restrictive about straight averaging. It is tantamount to stipulating that every world has the same logical weight in the evaluation of the closeness of a proposition to the actual world, an assumption which not only seems a little arbitrary but which in any case cannot be extended to infinite spaces. A theory of truthlikeness shouldn't be committed to a particular account of how much logical weight worlds have, and it certainly shouldn't rule out infinite spaces. The way to avoid these unwanted commitments would be to entertain different possible logical measures on the logical space. The third factor in the difference principle can then be generalized from *the number of worlds in A* to: *the logical weight of worlds u and v* . Thus amended, the difference principle, together with the uniform distance and pareto principles, entail that *weighted average distance* (or *expected distance*) is the only function to satisfy the distance extension principle.

Supervening on likeness between worlds is insufficient to characterize the likeness approach (since supervenient orderings that cut across the grain of likeness would also count), but it would be too restrictive to add the distance extension principles to the

²² Oddie (2013). The proof of this is a special case of a proof in Oddie and Milne (1991) to the effect that, given three parallel constraints, value is expected value of outcome.

characterization of the approach. Still, likeness/distance accounts that eschew weighted averaging must deny one or more of the three distance extension principles, all of which seem rather attractive.

There is, however, a way of sidestepping the extension problem altogether, by eschewing the “bottom-up” methodology that effectively generates it. This methodology begins by defining, or assuming, distances between finite worlds (or complete propositions) and then faces the problem of extending it both to complex worlds, and to arbitrarily large classes of such. There is, however, an alternative “top-down” methodology. That is, begin with general constraints on a theory of distance, or similarity, for suitable algebraic structures (Boolean algebras are the obvious candidates). Next, determine whether there exist distance functions satisfying those general constraints, and if so what their properties are. Further constraints could then be added to narrow down an adequate range of theories. On the top-down methodology it may or may not turn out that there exist qualifying similarity metrics that also deliver appealing results in simple or obvious cases. But one is spared the task of trying to define particular measures, where many of the basic choices—weightings of worlds, or of states or of properties—appear arbitrary. The top-down methodology has been explored quite thoroughly in papers by Miller, Mormann, and Gerla.

While a top-down methodology escapes both the extension and complexity problems, its weakness, at least as it has been developed so far, is its tendency to deliver a highly undesirable consequence, namely *VCF*. This may indicate that the initial constraints that have been chosen are defective, rather than that the methodology itself is defective.

23.7 THE COMPLEXITY PROBLEM

So far we have illustrated and developed accounts of truthlikeness within fairly simple, typically propositional, frameworks. Simplification is a legitimate methodological tool in philosophy as elsewhere, provided the simplifications are not essential. In the case of truthlikeness, we need evidence that the accounts can be extended to frameworks in which real theories are formulated.

Some of the features of the bottom-up approaches generalize fairly easily. For example, there is a straightforward generalization of weighted averaging from countable to uncountable logical spaces. Subject only to some modest measurability conditions we can show that there are three distance extension principles that entail that the distance of *A* from *w* is the average distance from *w* of worlds in *A*. But it is not at all clear that a measure of distance on worlds, based on the symmetric difference operation on atomic states, generalizes easily or even at all. A measure on the set of atomic states would provide one such way, but where the set of atomic states is infinite it is not clear whether there are plausible measures available.

One way of extending the measure of propositional distance to a more complex logical space involves cutting the space down to finite size, partitioning it into what are manageable, finite aspects of worlds relative to some parameters. One fruitful idea along these lines makes use of Hintikka's first-order *constituents* which, like the maximal conjunctions of a finite propositional framework, are jointly exhaustive, mutually exclusive, and finite.²³ Constituents lay out in a perspicuous manner all the different possible ways a collection of individuals can be related to each other, given a finite framework of attributes and given a finite parameter: *depth*. Every sentence in a first-order language comes with a certain quantificational depth—the number of embedded quantifiers. So for example *every electron is negatively charged* is depth-1; *every negatively charged particle repels any other* is depth-2; and *a helium atom consists of two negatively charged electrons and a positively charged nucleus* is depth-3. A proposition is *depth-d* if it has a first-order formulation at depth *d*. Hintikka showed that every depth-*d* proposition can be expressed by a finite disjunction of depth-*d* constituents, and that once inconsistent constituents are eliminated, the set of disjuncts for each proposition is unique.²⁴ This is the proposition's *normal form*. Constituents themselves are finite objects. They have a finite tree-structure, the nodes of which are analogous to conjunctions of atomic states. Consequently, if we can measure distance between constituents it is a simple matter to extend this to distance of depth-*d* propositions from the true depth-*d* constituent. The relative truthlikeness of propositions that cannot be characterized at any finite depth (i.e. non-finitely axiomatizable theories) can be characterized as some kind of limit, as depth is allowed to expand indefinitely.²⁵

Of course, first-order logic (the scope of Hintikka's normal form theorem) is itself a somewhat restrictive medium, one in which we cannot easily express all we would like to. Even simple claims about the fundamentality of particles are not felicitously expressed in a first-order idiom. That *electrons are fundamental* is not just the claim that every electron happens to have no constituents. Rather is the claim that the first-order property of *being an electron* has the second-order property of *being fundamental*. Necessarily, if *electronhood* is fundamental then every instance of electronhood has no constituent parts. A theory of truthlikeness ought to accommodate propositions only expressible in higher-order frameworks. (This has the added bonus of cutting down some non-finitely axiomatizable theories within the first-order realm to finitely axiomatizable size at the higher order.) To extend the normal form approach beyond the first order we need normal forms for higher-order logic, and happily it turns out there are such. A number of distance functions on constituents have been defined delivering a range of appealing results in fairly simple

²³ The idea was first suggested by Tichý at the end of his 1974, and worked out for first-order logic in his 1976. Niiniluoto independently worked out a version for monadic first-order logic in his 1977, a paper he had previously delivered at a conference in 1975.

²⁴ Hintikka (1963).

²⁵ Oddie (1978).

first-order cases, and (through a normal form theorem for higher-order languages) to higher-order cases as well.²⁶

Another aspect of complexity involves the fact that most interesting theories in science are not expressible in frameworks generated by basic properties and relations alone, but rather require continuous *magnitudes*—world-dependent functions from various entities to numbers. All the normal-form based accounts (even higher-order) are for systems that involve basic properties and relations, but exclude basic functions. There is, or course, a way of reducing n -argument functions to $n + 1$ -place relations, and then including real or complex numbers in the domain of lowest-level individuals. Not only is that a somewhat ad hoc solution to the problem, but no interesting applications based on such a reduction have been properly explored. It would clearly be preferable to give basic magnitudes the same status as basic properties in the generation of logical space. Niiniluoto has led the way in extending the likeness approach to functional spaces, but so far there is no good extension that both combines functions with properties and relations, and reduces the infinite complexity that these generate to finite or manageable proportions.²⁷ Until such an account is forthcoming, the likeness approach does not provide a comprehensive theory of closeness to truth. Of course, the infinite complexity problem is more acute for the bottom-up methodology and so, as with the extension problem, it is one that a top-down methodology could conceivably sidestep.

23.8 LANGUAGE-DEPENDENCE

The hard problem of truthlikeness involves language, or frame dependence.

Consider two states defined in terms of the familiar weather states: *minnesotan* =_{df} *hot* \Leftrightarrow *rainy*; and *arizonan* =_{df} *hot* \Leftrightarrow *windy*. The proposition $(h \wedge r \wedge w)$ is then tantamount to $(h \wedge m \wedge a)$; $(\neg h \wedge r \wedge w)$ to $(\neg h \wedge \neg m \wedge \neg a)$; and $(\neg h \wedge \neg r \wedge \neg w)$ to $(\neg h \wedge m \wedge a)$. In the *h-m-a*-framework the intuitive judgments of distances are reversed. Although this criticism was originally leveled at Tichý's version of the likeness approach,²⁸ it can be generalized and applies to any proposal (including versions of the relevant consequence approach) that delivers likeness-based intuitive judgments. If the argument is sound, then all the likeness-based intuitive judgments are themselves "language dependent."²⁹

It is tempting to say that *minnesotan* and *arizonan* are both "gerrymandered" or "derivative" conditions, in that their specifications involve reference to two distinct weather

²⁶ For theories of distance between constituents, see Niiniluoto (1977; 1987); Tichý (1976; 1978); Oddie (1986); Tuomela (1978).

²⁷ Niiniluoto (1987). Weston (1992) provides a semi-formal account of the notion of closeness of two models of a theory. See also Kieseppä (1996).

²⁸ Miller (1974; 1975a).

²⁹ See also Urbach (1983); Pearce (1983); Oddie (1986); Barnes (1990); Niiniluoto (1998); Weston (1992).

states. But the logical situation is formally symmetrical. Taking *minnesotan* and *arizonan* as primitive we can specify *rainy* and *windy* thus: *rainy* = $\text{df } \text{hot} \Leftrightarrow \text{minnesotan}$; *windy* = $\text{df } \text{hot} \Leftrightarrow \text{arizonan}$.

Despite the formal symmetry, one might resist the new conditions as less *natural* than our familiar weather states. States like *rainy* “carve reality and the joints” (in Plato’s memorable metaphor). *Rainy* is, and *minnesotan* is not, a *genuine universal*. But what makes a condition a *universal*? Either it would be a necessary and presumably a priori matter, or a contingent and presumably a posteriori matter, to be determined by mature science.³⁰ Either way, some properties are more fundamental, more basic, than others, and it is the fundamental properties, relations, and magnitudes that determine relations of distances between worlds.

A more radical response is to concede that nothing makes one class of conditions more primitive than another, but deny that the two frameworks yield genuine equivalences (Tichý 1976; 1978). Different primitives generate distinct possibilities, and consequently the propositions over those distinct logical spaces are themselves distinct entities.

There are arguments for this more radical position, but it saddles the realist with an unpalatable corollary: if the choice of primitive properties is not privileged by reality itself then those operating with different primitives cannot really communicate with one another, despite the fact the class of propositions to which they advert are correlated (in some sense) by necessity. However, what is clear is that for any theory of truthlikeness to replicate likeness-based or distance-based intuitions, a privileging of certain sets of conditions is essential.³¹ If no such privileging can be justified then not only are likeness-based theories inadequate, but so too are the intuitions which motivate them.

23.9 THE COMPATIBILITY OF THE APPROACHES

These different approaches are motivated by somewhat different desiderata. An interesting question that arises is whether, or to what extent, the approaches are compatible. If they are, then the different desiderata that motivate them might be accommodated in one happy theory. In their (2007), Zwart and Franssen offer precise, qualitative analyses of content and likeness orderings, and through an analogue of Arrow’s impossibility theorem, conclude that any attempt to meld content and likeness orderings will violate

³⁰ Armstrong (1978) is an extended argument for scientific realism about universals. Weston (1992) gives essentially this response to Miller. Barnes (1990) argues for a privileging of certain attributes based on certain epistemic features.

³¹ To see the broad reach of this translation-invariance argument, see Oddie (1986: ch. 6).

one or other of three basic desiderata. This move represents a welcome development in the debate between content and likeness theorists, lifting it above the fray of conflicting intuitions over particular cases into the realm of general desiderata.³² Schurz and Weingartner, however, fault Zwart and Franssen's analysis for essentially excluding the consequence approach altogether, thereby missing the possibility of compromise within an approach that incorporates both likeness and consequence desiderata. Further, their characterization of the two approaches rules Popper's account outside the content approach, and rules Tichý's account outside the likeness approach. So it is worth considering whether the three approaches are compatible under the characterization given here, which avoids both these shortcomings.

Call a distance/likeness function δ *trivial* if $\delta(v, w) = 1$ just in case $v \neq w$. A trivial distance function embodies an extreme version of *likeness nihilism*—namely, that as a matter of brute necessity no world is more like the actual world than is any other. It still counts as a possible view of likeness, albeit an extreme view, and one which is perhaps supported by a generalized language dependence argument. Given this trivial distance function, and weighted averaging, the distance of proposition A from the truth is $1 - (\mu(T)/\mu(A))$ if A is true, and 1 if A is false. Since this is generated by a distance function this measure of distance from the truth falls within the likeness approach, and since we used weighted averaging, it also satisfies the distance extension principles. Further, since the ordering both delivers *VCT*, and supervenes on truth and content, it falls within the content approach. It can also be shown to fall within the relevant consequence approach.³³ So, it turns out that the three approaches are, at least to this extent, compatible. Indeed Popper's ordering satisfies the strictures of all three approaches. It is obviously both a content and a consequence ordering, and since averaging the trivial distance function induces an extension of Popper's ordering, it falls within the likeness approach as well. Notice that the trivial distance measure, like Popper's original account, delivers both kinds of worthlessness of falsehoods: no falsehood is closer to the truth than any other, and no falsehood is closer to the truth than the least truthlike truth. And this is not just a peculiar feature of the trivial distance function for the result partially generalizes:

Any ordering of closeness to truth which is derived from averaging a likeness function, and which also delivers the value of content for truths, deems falsehoods absolutely worthless.³⁴

Although the three approaches are, strictly speaking, compatible, there is still a deep tension between them, which is illustrated by the following trilemma: you must either reject the distance extension principles, or reject the principle of the value of content for truths, or accept the absolute worthlessness of falsehoods.

³² Mormann (2005) puts this point very strongly.

³³ Oddie (2013).

³⁴ Oddie (2013).

23.10 NEW DIRECTIONS

Progress toward the truth in an inquiry can be achieved through a change not in what propositions are accepted or rejected, but rather in the distribution of credences over propositions. The probabilistic epistemic utility program, like the propositional truthlikeness program, is also an attempt to pin down the notion of accuracy—that of probabilistic credal states.

The target credal state, P^T , is the one that assigns 1 to the true hypothesis and zero to its rivals, but other credal states are closer to the target than others. Suppose P_0 distributes probabilities equally over the eight weather states. In comes some new evidence, relayed from various weather stations say, and while it doesn't rule out any of the eight possibilities definitively, it bumps the probability of $h \wedge r \wedge w$ (the truth as it happens) up to 0.3, and reduces the probability of its rivals to 0.1 (Table 23.1).

The move from P_0 to P_1 is a move in the direction of P^T , even though no new proposition is fully accepted. Now suppose a bit more evidence comes in and the probabilities are updated to P_2 . The move from P_1 to P_2 uniformly shifts probabilities from less accurate to more accurate states. P_2 is closer to P^T than is P_1 , which in turn is closer than P_0 . The idea that probabilistic credal states have degrees of cognitive value proportional to their closeness to the target credal state has proved very fruitful. Provided the cognitive value function satisfies an apparently modest constraint known as *Propriety*, one can vindicate many of the core tenets of a Bayesian epistemology:³⁵ for example, that credences obey the standard axioms of probability;³⁶ that updating should go by conditionalization;³⁷ that experimenting has positive expected cognitive value;³⁸ and that credences and propensities obey the Principal Principle.³⁹

	$(h \wedge r \wedge w)$	$(h \wedge r \wedge \neg w)$	$(h \wedge \neg r \wedge w)$	$(\neg h \wedge r \wedge w)$	$(h \wedge \neg r \wedge \neg w)$	$(\neg h \wedge r \wedge \neg w)$	$(\neg h \wedge \neg r \wedge w)$	$(\neg h \wedge \neg r \wedge \neg w)$
P_0	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125
P_1	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1
P_2	0.315	0.17	0.17	0.17	0.05	0.05	0.05	0.0025
P^T	1	0	0	0	0	0	0	0

TABLE 23.1 Credal states in the weather framework

³⁵ *Propriety* says, roughly, that each credal state judges that it itself maximizes expected cognitive value.

³⁶ Joyce (1998; 2009); Leitgeb and Pettigrew (2011).

³⁷ Oddie (1997); Greaves and Wallace (2007).

³⁸ Oddie (1997); Fallis (2007); Myrvold (2012).

³⁹ Pettigrew (2013).

Since both the truthlikeness program and the epistemic utility program are zeroing in on accuracy, consilience should be in the offing. Unfortunately a compelling principle governing truthlikeness—*Proximity*—appears to be incompatible with what is perhaps the central principle of epistemic utility theory—namely, *Propriety*.⁴⁰ The implications of this incompatibility have not yet been fully worked out.

There are other problems of probability and acceptance for which the notion of truthlikeness may well provide interesting new insights. Among these are belief revision, the preface paradox, and the lottery paradox.⁴¹ Further work in these areas should prove illuminating.

23.11 THE VALUE OF TRUTH AND THE VALUE OF TRUTHLIKENESS

We began with the presumption that truth is a cognitive value, but we can now see that it is not the only veritistic cognitive value. Accuracy, or closeness to the truth, should also be valued. It is better, other things being equal, to embrace a proposition that is close to the truth than one that is far away. Moreover, if a falsehood may be very close to the truth, while a truth may lie some distance away, then *being true* is, *in itself*, of overridable cognitive value. Indeed, on any of the three approaches to truthlikeness that we have considered, it is possible for a falsehood to be closer to the whole truth of some matter than its own (true) negation. If this is right, it may not always be better to believe a true proposition than its false negation. So an interesting corollary of the study of truthlikeness is that it undermines the ancient hegemony of the true/false dichotomy.

REFERENCES

- Armstrong, D. M. (1978). *Universals and Scientific Realism*. Cambridge: Cambridge University Press.
- Barnes, E. (1990). "The Language Dependence of Accuracy," *Synthese* 84: 59–95.
- Brink, C. and J. Heidema (1987). "A Verisimilar Ordering of Theories Phrased in a Propositional Language," *British Journal for the Philosophy of Science* 38: 533–49.
- Burger, I. C. and J. Heidema (1994). "Comparing Theories by Their Positive and Negative Contents," *British Journal for the Philosophy of Science* 45: 605–30.
- Cevolani, G. (2017). "Fallibilism, Verisimilitude, and the Preface Paradox," *Erkenntnis*, 82:169–83.

⁴⁰ See Oddie (2017). *Proximity* says, roughly, that one cannot decrease cognitive value by shifting probabilities to states closest to the truth; e.g. the standard measure of epistemic utility—the *Brier score*—deems P_2 more inaccurate than P_1 , which contradicts *Proximity*. Other EU measures satisfying *Propriety* can capture this particular ordering, but none can capture all the intuitively desirable orderings.

⁴¹ See Cevolani, Crupi, and Festa (2011); Cevolani and Schurz (2017); Cevolani (2017).

- Cevolani, G. and G. Schurz (2017). "Probability, Approximate Truth, and Truthlikeness: More Ways out of the Preface Paradox," *Australasian Journal of Philosophy*, 95: 209–25.
- Cevolani, G., V. Crupi, and R. Festa (2011). "Verisimilitude and Belief Change for Conjunctive Theories," *Erkenntnis* 75: 183–202.
- Fallis, D. (2007). "Attitudes Toward Epistemic Risk and the Value of Experiments," *Studia Logica* 86: 215–46.
- Gemes, K. (2007). "Verisimilitude and Content," *Synthese* 154: 293–306. doi:10.1007/s11229-005-2565-x.
- Gerla, G. (2007). "Point-Free Geometry and Verisimilitude of Theories," *Journal of Philosophical Logic* 36: 707–33.
- Greaves, H. and D. Wallace (2006). "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility," *Mind* 115: 607–32.
- Hilpinen, R. (1976). "Approximate Truth and Truthlikeness," in J. Przelecki et al. (eds.), *Formal Methods in the Methodology of the Empirical Sciences*. Dordrecht: Reidel, 19–42.
- Hintikka, J. (1963). "Distributive Normal Forms in First-Order Logic," in J. N. Crossley et al. (eds.), *Formal Systems and Recursive Functions*. Amsterdam: North-Holland, 47–90.
- Joyce, J. M. (1998). "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science* 65: 575–603.
- Joyce, J. (2009). "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief," in F. Huber and C. Schmidt-Petri (eds.), *Degrees of Belief*, Synthese Library Series 342. Dordrecht: Springer, 263–97.
- Kieseppä, I. (1996). *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*. Dordrecht: Kluwer.
- Krantz, D. H., R. D. Luce, P. Suppes, and A. Tarski (1971). *Foundations of Measurement*, vol. 1. New York: Academic Press.
- Kuipers, T. (1987). "A Structuralist Approach to Truthlikeness," in T. A. Kuipers (ed.) *What is Closer-to-the-Truth? A Parade of Approaches to Truthlikeness*. Amsterdam: Rodopi, 79–99.
- Kuipers, T. (1992). "Naïve and Refined Truth Approximation," *Synthese* 93: 299–342.
- Kuipers, T. (2000). *From Instrumentalism to Constructive Realism*. Dordrecht: Kluwer.
- Leitgeb, H. and R. Pettigrew (2010). "An Objective Justification of Bayesianism I: Measuring Inaccuracy," *Philosophy of Science* 77: 201–35.
- Miller, D. (1974). "Popper's Qualitative Theory of Verisimilitude," *British Journal for the Philosophy of Science* 25: 166–77.
- Miller, D. (1977). "On Distance from the Truth as a True Distance," *Bulletin of the Section of Logic* 6: 15–23.
- Miller, D. (2007). *Out Of Error: Further Essays on Critical Rationalism*. Aldershot: Ashgate.
- Mortensen, C. (1983). "Relevance and Verisimilitude," *Synthese* 55: 353–65.
- Mormann, T. (2005). "Geometry of Logic and Truth Approximation," *Poznań Studies in the Philosophy of the Sciences and the Humanities* 83: 431–54.
- Mormann, T. (2006). "Truthlikeness for Theories on Countable Languages," in *Karl Popper: A Centenary Assessment*, vol. 3, ed. I. Jarvie, K. Milford, and D. Miller. Aldershot: Ashgate, 3–15.
- Myrvold, W. C. (2012). "Epistemic Values and the Value of Learning," *Synthese* 187: 547–68.
- Niiniluoto, I. (1977). "On the Truthlikeness of Generalizations," in R.E. Butts et al. (eds.), *Basic Problems in Methodology and Linguistics*. Dordrecht: Reidel, 121–47.
- Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.
- Niiniluoto, I. (1998). "Verisimilitude: The Third Period," *British Journal for the Philosophy of Science* 49: 1–29.

- Oddie, G. (1978). "Verisimilitude and Distance in Logical Space," *Acta Philosophica Fennica* 30: 227–43.
- Oddie, G. (1986). *Likeness to Truth*. Dordrecht: Reidel.
- Oddie, G. (1997). "Conditionalization, Cogency, and Cognitive Value," *British Journal for the Philosophy of Science* 48: 533–41.
- Oddie, G. (2013). "The Content, Consequence and Likeness Approaches to Verisimilitude: Compatibility, Trivialization, and Underdetermination," *Synthese* 190: 1647–87.
- Oddie, G. (2017). "What Accuracy Could Not Be," *British Journal for the Philosophy of Science*, axx032, <https://doi.org/10.1093/bjps/axx032>. Published online: 1 September 2017.
- Oddie, G. and P. Milne (1991). "Act and Value: Expectation and the Representability of Moral Theories," *Theoria* 57: 42–76.
- Pearce, D. (1983). "Truthlikeness and Translation: a Comment on Oddie," *British Journal for the Philosophy of Science* 34: 380–5.
- Pettigrew, R. (2011). "Epistemic Utility Arguments for Probabilism," in E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), <http://plato.stanford.edu/archives/win2011/entries/epistemic-utility/>
- Pettigrew, R. (2013). "A New Epistemic Utility Argument for the Principal Principle," *Episteme* 10: 19–35.
- Popper, K. R. (1963). *Conjectures and Refutations*. London: Routledge.
- Popper, K. (1976). "A Note on Verisimilitude," *British Journal for the Philosophy of Science* 27: 147–59.
- Schurz, G. and P. Weingartner (1987). "Verisimilitude Defined by Relevant Consequence-Elements," in T. A. Kuipers (ed.), *What is Closer-to-the-Truth? A Parade of Approaches to Truthlikeness*. Amsterdam: Rodopi, 47–78.
- Schurz, G. and P. Weingartner (2010). "Zwart and Franssen's Impossibility Theorem Holds for Possible-World-Accounts but Not for Consequence-Accounts to Verisimilitude," *Synthese* 172: 415–36.
- Smith, N. (2005). "Vagueness as Closeness," *Australasian Journal of Philosophy* 83: 157–83.
- Tichý, P. (1974). "On Popper's Definitions of Verisimilitude," *British Journal for the Philosophy of Science* 25: 155–60.
- Tichý, P. (1976). "Verisimilitude Redefined," *British Journal for the Philosophy of Science* 27: 25–42.
- Tichý, P. (1978). "Verisimilitude Revisited," *Synthese* 38: 175–96.
- Tuomela, R. (1978). "Verisimilitude and Theory Distance," *Synthese* 38: 213–46.
- Urbach, P. (1983). "Intimations of Similarity: the Shaky Basis of Verisimilitude," *British Journal for the Philosophy of Science* 34: 166–75.
- Weston, T. (1992). "Approximate Truth and Scientific Realism," *Philosophy of Science* 59: 53–74.
- Zwart, S. D. (2001). *Refined Verisimilitude*. Dordrecht: Kluwer.
- Zwart, S. D. and M. Franssen (2007). "An Impossibility Theorem for Verisimilitude," *Synthese* 158: 75–92.

CHAPTER 24

TRUTH IN MATHEMATICS

ØYSTEIN LINNEBO

24.1 INTRODUCTION

MANY questions in the philosophy of mathematics are concerned with the nature and role of the concept of truth in mathematics. I shall here be concerned with four such questions.

One question is whether the concept of truth is needed in a philosophical account of mathematics. Do we need to attribute truth and falsity to mathematical statements in the way that we do to empirical statements? Quite clearly, mathematics operates with strict standards of correctness. For instance, philosophers of every orientation agree that it is correct to say that $7 + 5 = 12$ and incorrect to say that $7 + 5 = 13$. But it is not obvious that this notion of correctness needs to be identified with truth. Many philosophers have argued that the notion of correctness that is operative in mathematics falls short of full-fledged truth and that this notion is better understood in terms of acceptability by certain agreed standards. I shall argue that such arguments should be resisted and that the concept of truth is indeed needed in mathematics.

Another question concerns the nature of mathematical truth. On a classical conception of truth, the truth of a mathematical sentence requires the existence of mathematical objects for its singular terms to denote and its quantifier to range over. This forms the core of an influential argument developed by Frege for the existence of abstract mathematical objects. In order to avoid this ontology of abstract objects and the philosophical puzzles to which it gives rise, other philosophers have proposed various non-classical conceptions of mathematical truth, which allow mathematical sentences to be truth without there being any mathematical objects. I shall argue that these non-classical conceptions are problematic, whereas the classical conception enjoys substantial support.

A third question concerns the relation between the existence of mathematical objects and the objectivity of mathematical truth. According to a traditional platonist view, the former explains the latter. It is because there exist mind-independent

mathematical objects with mind-independent properties that mathematical sentences have their objective truth-values. I discuss some considerations, due to Frege and others, that favor the opposite direction of explanation: from mathematical truth to mathematical objects.

The final question is the extent to which the truth-values of mathematical statements are objective. Do mathematical statements have their truth-values independently of our current choice of axioms and our ability to prove theorems that follow from these axioms? This question has great significance for the methodology of mathematics. The search for new axioms for some branch of mathematics is typically motivated by the belief that some of the statements that are left undecided by our current axioms have an objective truth-value that can and should be made explicit by the adoption of further axioms.

24.2 WHY MATHEMATICS NEEDS A CONCEPT OF TRUTH

Our first question is whether a concept of truth is needed in a philosophical account of mathematics. Can we make sense of mathematical language and practice without invoking the notions of truth and falsity?

Let's begin with the slightly narrower question of whether the notion of correctness that governs mathematical practice needs to be identified with the notion of truth. In other areas of discourse, there are good reasons to take truth to be the principal notion of theoretical correctness. Set aside practical notions of correctness that are concerned with the demands of instrumental rationality, ethics, or etiquette. Then it seems that a statement about, say, the furniture in my room is correct just in case it is true. Why should mathematics be an exception to this identification of theoretical correctness with truth? Much of the resistance to this identification stems from the difficulty of making sense of what would make a mathematical statement true or false. A statement about my furniture is made true by my furniture (or lack thereof) and its properties. But it is far less obvious what would make a mathematical statement true. The platonist's answer that mathematical statements are made true by an abstract realm of mathematical objects seems needlessly speculative. Many philosophers have found it more prudent to analyze mathematical discourse and practice in terms of notions that are less problematic. Among the least problematic facts about mathematics is that mathematicians prove theorems and that there is wide agreement in the mathematical community about the standards for assessing the correctness of proofs. So why not try to use these relatively unproblematic facts as a starting point for our philosophical analysis?

These reflections motivate a *formalist* approach to mathematics, according to which the highest notion of correctness that is found in mathematics, or is needed in its analysis, is theoremhood in some axiomatic theory. For instance, on this view the

correctness of the statement that there are infinitely many prime numbers consists in its being a theorem of the relevant axiomatic theory, for instance Peano arithmetic.

An extreme version of this view is *game formalism*, which regards mathematics as a meaningless game with symbols.¹ The axioms describe the strings of symbols that count as legitimate “initial positions,” and the rules of inference characterize the permissible “moves” that allow one to progress from some strings of symbols to another. A mathematical theory is thus much like the game of chess, except that its moves are made with strings of symbols rather than pieces on a board. In particular, a string of mathematical symbols is just as devoid of meaning as a configuration of chess pieces. So such a string admits of truth or falsity no more than a configuration of chess pieces.

There are milder versions of formalism as well. Instead of denying that mathematical language has any sort of meaning, one can limit oneself to the weaker claim that all philosophically important notions can be analyzed in terms of purely syntactic properties of mathematical language, making no appeal to any semantic properties. This weaker claim can also be combined with a liberal view of what counts as a syntactic property.²

Probably the most famous defense of the need for a notion of truth in mathematics is due to Frege, who writes as follows:

Why can arithmetical equations be applied? Only because they express thoughts. How could we possibly apply an equation which expressed nothing and was nothing more than a group of figures, to be transformed into another group of figures in accordance with certain rules? Now it is applicability alone which elevates arithmetic from a game to the rank of a science. (Frege 1903: section 91; trans. in Frege 1952)

Unlike a configuration of chess pieces, the string of symbols that make up a mathematical theorem says something. And it is because what the theorem says is true that it is applicable in scientific arguments and explanations.

In order to explain and evaluate Frege’s argument, it will be useful to consider an example of how pure mathematics can contribute to a scientific explanation. So let’s consider Euler’s famous problem of the Königsberg bridges. The great eighteenth-century mathematician Euler wondered whether it is possible to choose a route through the city of Königsberg that involves crossing each of its seven bridges once and only once. He showed how the problem can be formulated in the abstract terms of what is now known as graph theory. Each land mass can be represented by a single node, and each bridge by an edge connecting two nodes. A simple mathematical analysis of the resulting graph then demonstrates that, given the arrangement of rivers and bridges in Königsberg, there can be no route of the desired kind. This establishes the empirical claim that no

¹ See Resnik (1980) and Shapiro (2000b) for discussion.

² An example of a liberal view of this sort is the program canvassed in Carnap (1934), where for instance the omega-rule is counted as syntactic. (This is the infinitary arithmetical rule which says that, from premises $\phi(\bar{n})$ for every natural number n , one may infer $\forall n\phi(n)$.)

matter what route Euler chooses through the city, he will not manage to cross each bridge precisely once. How does Euler's mathematical argument succeed in assuring us of the truth of this empirical conclusion? The most obvious answer is that this is accomplished because the argument is sound: it is a logically valid argument from true premises. But this answer presupposes a notion of truth that is applicable to empirical and mathematical premises alike, and which is such that the relevant theorem of graph theory does indeed count as true.

An interesting objection to Frege's argument is suggested by the work of Hartry Field, using ideas that go back to the great mathematician David Hilbert.³ Consider a valid argument whose premises are either theorems of pure mathematics or nominalistic truths about the physical world, and whose conclusion C is another nominalistic statement. In order to explain why the conclusion C is guaranteed to be true, there is no need to ascribe truth to the mathematical theorems, as Frege's explanation does. It suffices that the mathematical theorems have a weaker property known as *conservativeness*. Let T_1 be a theory about the physical world formulated in a nominalistic language \mathcal{L}_1 . Let \mathcal{L}_2 be a language that extends \mathcal{L}_1 by adding some mathematical vocabulary, and let T_2 be a theory in \mathcal{L}_2 that extends T_1 . Then T_2 is said to be conservative over T_1 just in case every consequence of T_2 that is formulated in the restricted language \mathcal{L}_1 is also a consequence of T_1 . Field then observes that in order to establish that the conclusion C is true, it suffices to show that the mathematical premises are conservative over the nominalistic statements.

However, a lot of hard work remains before this strategy will enable us to dispense with all ascriptions of truth to mathematical statements. First, we would have to provide a purely nominalistic formulation of all of science. Field shows how this can be done for Newtonian gravitational theory, which is a good start. But it remains uncertain whether all of contemporary science can be nominalized in this way.⁴

Second, the conservativeness (and thus also consistency) of the various theories of pure mathematics would have to be established. And this would have to be done in a way that avoids ascribing truth to these theories. Here another observation inspired by Frege is relevant. Sometimes our best evidence for the consistency of a theory is also evidence for the truth of the theory.⁵ Consider for instance our conviction that ZFC set theory is consistent. This conviction is not based primarily on the empirical fact that no contradiction has yet been found. Rather it is based on our having a reasonably clear grasp of the intended model for ZFC, namely the iterative hierarchy of sets. Our conviction that ZFC is consistent rests on the observation that its axioms are true in this intended model.

³ See respectively Field (1984) and Hilbert (1926). Of course, Field is not a formalist but an error theorist: he accepts that mathematical statements have truth-values but takes all atomic such statements to be false. But formalists and error theorists are united in their opposition to Frege's argument.

⁴ See Burgess and Rosen (1997: part II).

⁵ See Frege (1903: section 144 (repr. in Frege 1952)), as well as the Frege–Hilbert correspondence, repr. in Frege (1984). See Field (1991) for a defense of the opposite view.

A third worry concerns the fact that Field's project is a reconstructive one. Science isn't actually done in the way that Field suggests. Our best scientific theories aren't formulated in a nominalistic language but freely make use of large amounts of mathematics. So even if the first two worries could be addressed, the challenge would still remain of making sense of actual scientific practice.⁶ How do actual scientific arguments, with their liberal mix of mathematical and physical vocabularies and considerations, succeed in establishing true conclusions? Here Frege's argument still has force and gives us a reason to ascribe truth to premises formulated in a vocabulary that is partly or wholly mathematical.

The third worry points to another argument for the truth of mathematical theorems, which has received much attention in recent years. This argument can be formulated as follows.

- P1. Mathematicians' attitude toward mathematical theorems is largely one of taking-to-be-true.
- P2. Mathematicians' attitude toward mathematical theorems is largely appropriate.
- C. Mathematical theorems are by and large true.

The argument is clearly valid, since the attitude of taking-to-be-true is appropriately taken only toward claims that are in fact true. But are the premises true?

The second premise, P2, is often defended by observing that mathematics is an extremely successful science, concerning which mathematicians possess great expertise. So there is reason to believe that whatever attitude mathematicians take to their theorems is indeed appropriate. This argument is often given a "naturalistic" development, to the effect that philosophers should defer to the opinions of working scientists, whose track record is better than that of philosophers.⁷ Although this form of naturalism would certainly suffice to support P2, it is not necessary. Many philosophers who are deeply opposed to this form of naturalism nevertheless believe that the scientific legitimacy of mathematics should serve as a datum for philosophical theorizing and not be explained away. Kant and Frege are two famous examples. This philosophical orientation too supports P2.

The prospects for denying P1 seem better. The task of determining what attitude mathematicians take to their theorems falls to philosophers or psychologists, not to mathematicians themselves. So here it would be inappropriate simply to defer to mathematicians' opinion. Some philosophers have objected to P1 by invoking a form of *fictionalism*, according to which mathematicians' attitude toward mathematical theorems is some form of acceptance that falls short of taking-to-be-true. For instance, mathematicians accept the claim that there are infinitely many primes but don't actually take it to be true.

⁶ Cf. Burgess and Rosen (1997) and Maddy (1997).

⁷ See Burgess and Rosen (1997); Lewis (1991: 57–9); and Maddy (1997).

However, this objection to P1 is problematic. Is there really room for two kinds of favorable attitude—acceptance and taking-to-be-true—to live side by side? There is no phenomenological evidence that mathematicians’ acceptance of a mathematical theorem is any less wholehearted than that involved in their taking various empirical claims to be true. Nor does the view enjoy any solid psychological evidence detectable from an objective, third-person point of view. For instance, mathematicians are not any more disposed to retract their claims when challenged than physicists are. This stands in stark contrast to cases that do involve two kinds of favorable attitude. For instance, meteorologists often find it convenient to talk about Coriolis forces, although they know perfectly well that such forces are purely fictional and invoked only as an innocent trick to simplify arguments and calculations. In such cases meteorologists are consciously aware that their “acceptance” of the relevant claims falls short of a commitment to their truth. And this awareness is externally detectable as a tendency to retract or reformulate the relevant claims when challenged.

I conclude that a strong case can be made that the notion of truth is indeed needed in mathematics. However, as it stands, this argument places few or no constraints on the nature of the notion of truth that is found in mathematics. For all we know so far, this notion may be very minimal or otherwise very different from the notion of truth that is applicable to ordinary empirical statements. To investigate this, I turn in the next section to the question of the semantics for mathematical language.

24.3 DIFFERENT CONCEPTIONS OF MATHEMATICAL TRUTH

In light of the discussion in the previous section, I shall henceforth make the following assumption:

Mathematical Truth. Most sentences accepted as mathematical theorems are true (regardless of their syntactic and semantic structure).

So far, this is not very surprising. Successful sciences discover truths, and mathematics is no exception. But in 1884, Frege observed that this assumption, when combined with another assumption about the correct semantics for the language of mathematics, has the surprising implication that there exist mathematical objects (Frege 1953). I shall refer to this as the *Fregean argument*.

The second assumption can be formulated as follows.

Classical semantics. The singular terms of the language of mathematics purport to refer to mathematical objects, and its first-order quantifiers purport to range over such objects.

The word “purport” needs to be explained. When a sentence S purports to refer or quantify in a certain way, this means that for S to be true, S must succeed in referring or quantifying in this way.

The Fregean argument goes as follows. Consider sentences that are accepted as mathematical theorems and that contain one or more mathematical singular terms. By Mathematical Truth, most of these sentences are true.⁸ Let S be one such sentence. By classical semantics, the truth of S requires that its singular terms succeed in referring to mathematical objects. Hence there must be mathematical objects.

In fact, if there are any mathematical objects, they would presumably be abstract (i.e. non-spatiotemporal and causally inefficacious); otherwise mathematicians’ attitude toward such objects would be thoroughly misguided.⁹ This makes the following conclusion appropriate:

Anti-nominalism. There are abstract mathematical objects.

Let’s now take a closer look at the assumption of classical semantics. The claim is that the language of mathematics functions semantically much like language in general has traditionally been assumed to function; that is, the semantic function of singular terms and quantifiers is respectively to refer to objects and to range over objects. This is a broadly empirical claim about the workings of a semi-formal language used by the community of professional mathematicians.¹⁰ Note also that classical semantics is compatible with most traditional views on semantics. In particular, it is compatible with all the standard views on the meanings of sentences, namely that they are truth-values, propositions, or sets of possible worlds.

Classical semantics enjoys strong *prima facie* plausibility. For the language of mathematics certainly appears to have the same semantic structure as ordinary non-mathematical language.¹¹ As Burgess (1999: 288) observes, the following two sentences appear to have the same simple semantic structure of a predicate being ascribed to a subject:

- (1) Evelyn is prim
- (2) Eleven is prime

This appearance is also borne out by the semantic theories developed by linguists and semanticists. These theories do not exempt mathematical language from the general semantic analyses that they propose.

⁸ Note that this step uses the parenthetical precisification in Mathematical Truth, without which it would be possible for most sentences accepted as mathematical theorems to be true yet all sentences of the form mentioned in the text to be false.

⁹ For critical discussion of the distinction between concrete and abstract, see Rosen (2017).

¹⁰ In the widely adopted terminology of Burgess and Rosen (1997: 6–7), Classical semantics is a *hermeneutic* claim; i.e. it is a descriptive claim about how a certain language is actually used, not a normative claim about how this language ought to be used.

¹¹ This is why Benacerraf (1973) argues there is a strong presumption in favor of a unified semantics, which treats mathematical and non-mathematical language in analogous ways.

However, since the philosophical stakes are high, it is unsurprising that classical semantics has been challenged. Perhaps the apparent similarities between mathematical and non-mathematical language are deceptive and mathematical language would be better analyzed in some alternative way. I shall now consider four non-classical conceptions of how mathematical sentences get their truth-values.

The first non-classical conception seeks to reduce mathematical truth to theoremhood. This is an example of what Benacerraf (1973) calls a “combinatorial” account of truth. The idea is that a mathematical sentence is true just in case it is a theorem—in some appropriate sense. But what is the appropriate notion of theoremhood?

A notion of theoremhood that is tied to some fixed formal system S is unlikely to be appropriate for the present purpose. To see this, note first that any useful formal system has to be recursively axiomatizable. Gödel’s incompleteness theorems thus imply that, if S is consistent and rich enough to express a certain amount of arithmetic, it will be incomplete; in particular, S will be unable to prove the arithmetical statement $\text{Con}(S)$ which expresses that S is consistent. However, acceptance of a formal system S implicitly involves acceptance of its consistency and thus also of $\text{Con}(S)$. For one would never accept a formal system unless one thought it was consistent.

This informal argument can be made technically more precise as follows. Let S' be the formal system that results from S by adding variables and quantifiers of one order higher than those found in S , as well as the ordinary rules and axioms governing these quantifiers. We can then define in S' a compositional truth predicate applicable to all sentences of S .¹² This enables us to give a proof in S' of $\text{Con}(S)$, something we could not do in the weaker system S . Since it is unreasonable to accept the standards of proof associated with S without also being willing to accept those associated with S' , this shows that it is unreasonable to tie mathematical truth to theoremhood in S or any other fixed formal system.

So if mathematical truth is to be reduced to theoremhood, then the relevant notion of theoremhood should not be tied to some fixed formal system. As it turns out, mathematicians often speak about proof without having any specific formal system in mind. A proof in this informal sense is not tied to a fixed formal system but is seen as a sound mathematical argument whose assumptions can be recognized as acceptable axioms by all competent mathematicians. Perhaps mathematical truth can be reduced to the corresponding, informal notion of theoremhood. However, some serious difficulties remain. The informal notion of provability is poorly understood.¹³ Moreover, no satisfactory proof-based semantics has been developed for the language of mathematics, despite several decades of attempts by philosophers and logicians such as Dummett and Prawitz.¹⁴

¹² Alternatively, one could add such a truth-predicate as a new primitive symbol, governed by the appropriate axioms.

¹³ See, however, Leitgeb (2009).

¹⁴ See Dummett (1991) and Prawitz (2006), as well as the other contributions to the special issue of *Synthese* where the latter appears.

A second non-classical semantics is inspired by one of the traditional formalist approaches to mathematics. Let's consider the case of arithmetic, where this approach has its greatest plausibility. The semantics in question denies that the numerals play any direct referential role, claiming instead that they function as symbols on which computations are carried out in accordance with certain "rewrite rules." For instance, one such rule says that $m + S(n)$ may be rewritten as $S(m + n)$; there are other rules governing the other symbols. An arithmetical equation is then said to be true just in case it can be turned into a tautology by means of these rewrite rules.¹⁵

One problem with this semantics concerns the quantifiers. How are they to be analyzed? Perhaps the most natural approach is to analyze the quantifiers as infinite conjunctions and disjunctions. However, this involves an extreme idealization of what mathematicians and ordinary people mean by their quantifiers. A semantics based on such an extreme idealization effectively gives up on one of the main goals of semantic theorizing, namely to explain the semantic competence that underlies people's understanding of the language in question. Another problem is that the proposed semantics adopts a non-standard interpretation of the identity predicate "=". Ordinarily, the identity predicate is concerned with the identity of the referents of the terms flanking it; but in mathematics, the predicate is, on the present view, concerned with these terms themselves. This dichotomy is unnatural and poorly motivated. In fact, the proposed semantics seems to conflate the meaning of the identity predicate with our procedures for verifying identity facts.

A third non-classical semantics is associated with the *modal structuralism* proposed in (Putnam 1967) and developed in (Hellman 1989). Let's again consider the case of arithmetic. Modal structuralists reject the platonist view that the language of arithmetic is concerned with a particular system of abstract natural numbers. Rather, they understand arithmetical sentences as making assertions about what is necessarily the case in any system of objects that are structured in the way that the natural numbers are supposed to be structured. This structure is described up to isomorphism by the axioms of second-order Peano arithmetic. Let PA2 be the conjunction of these axioms. Then $\text{PA2}(X, f, a)$ expresses the claim that the relevant structure is instantiated by the collection X of objects (playing the role of the natural numbers), the function f (playing the role of the successor function that maps one number n to its immediate successor $n + 1$), and the designated object a (playing the role of zero). The modal structuralist analysis of an arithmetical sentence A can now be formalized as follows:

$$(1) \quad \Box \forall X \forall f \forall a [\text{PA2}(X, f, a) \rightarrow A(X, f, a)]$$

¹⁵ This semantics departs from, but improves on, the view known as *term formalism*, which holds that the numerals denote themselves, understood as either types or tokens. This ascription of reference has bizarre and unintended consequences, such as the truth of sentences such as "o has topological genus one" and "1 is nearly vertical."

Is this an adequate account of arithmetical language as used by professional mathematicians and competent lay people? It will be objected that neither mathematicians nor lay people have in mind these sorts of modalized generalizations when they use arithmetical language. Although I believe this objection has force, it is not obviously conclusive. For instance, Pettigrew (2008) proposes that the ordinary symbols for the collection of natural numbers, the successor function, and zero can be understood as “designated free variables,” whose values are implicitly assumed to satisfy the antecedent of (1). This provides the resources for a significant response to the above objection.

The final non-classical account of how mathematical sentences get their truth-values is a version of fictionalism. Where the previous version of fictionalism was based on a non-standard view of *the attitude* that mathematicians take to their utterances, this version is based on a non-standard view of *the content* of these utterances. The idea is that in mathematical discourse, language is used in some non-literal way. For instance, Stephen Yablo argues that, although the literal content of a mathematical sentence often involves a variety of exotic mathematical objects, its “real content” is often just that the world is in a particular kind of state, namely a state that makes it appropriate to make the relevant mathematical assertion.¹⁶ For instance, the real content of “the number of planets is 8” is just that there are eight planets. Mathematical objects are invoked primarily because they provide an easier way of conveying the real content. For instance, the real content of “the number of planets is prime” could not be conveyed as easily without talking about mathematical objects.

This version of fictionalism has faced a variety of objections.¹⁷ I shall mention two. The first objection concerns systematicity. Non-literal language use is often quite unsystematic. For instance, metaphors can be highly creative, and the metaphorical effect is fragile and can be destroyed by the replacement of an expression with another with the same semantic value. By contrast, the language of mathematics is one of the more systematic parts of natural language. So *prima facie*, the comparison of mathematical language with metaphorical language is implausible. However, systematicity need not pose an insuperable problem for all non-literal approaches to mathematical language. Although metaphors and many other forms of non-literal language use are highly unsystematic, there is nothing inherent in the idea of non-literal use that conflicts with systematicity. Just as there are systematic rules governing literal meaning, there could in principle be languages with a parallel set of systematic rules governing non-literal meaning.

This brings us to the second objection. Is there really room for two kinds of systematic meaning—literal and non-literal—to live side by side? There is no phenomenological evidence that in mathematics, language is being used in a special, non-literal way. When I say that there are infinitely many primes, it seems to me that I mean precisely what my

¹⁶ See Yablo (2005). However, as Yablo is acutely aware, the specification of the real content is often a delicate matter; see Yablo (2001).

¹⁷ See Stanley (2001).

words mean. Nor is there any robust psychological evidence available from an objective, third-person point of view to support the view that mathematical language tends to be used in some non-literal way. For instance, mathematicians are no more disposed to paraphrase or restate their claims when challenged than other scientists are. By contrast, such evidence is not hard to come by in paradigmatic cases of non-literal language use. For instance, when language is used metaphorically, speakers tend to be aware that certain words' literal meanings are being exploited to create a special, non-literal effect. For this reason, they are also disposed to paraphrase or explain their meaning when challenged.

I conclude that classical semantics is so called with good reason, as there is substantial evidence in favor of it, and that the non-classical alternatives face serious problems.

24.4 OBJECTS AND OBJECTIVITY IN MATHEMATICS

I shall now describe the two forms of realism about mathematics and examine the relation between them. One form is concerned with the independent existence of mathematical objects, and the other, with the objectivity of mathematical truths.

The first form of realism can be expressed as follows.

Mathematical platonism. There are abstract mathematical objects, and these objects exist independently of intelligent agents and their language, thought, and practices.

Mathematical platonism goes beyond the thesis of anti-nominalism, which we encountered in section 24.3. For the latter thesis says just that there are abstract mathematical objects, whereas the former adds a claim about the independent existence of these objects. This independence claim is meant to cash out an analogy between mathematics and the natural sciences. Just as electrons and planets exist independently of us, so do numbers and sets. And just as statements about electrons and planets are made true or false by the objects with which they are concerned and these objects' perfectly objective properties, so are statements about numbers and sets. Mathematical truths are therefore discovered, not invented.

The second form of realism about mathematics can be formulated as follows.

Truth-value realism. Every well-formed mathematical statement has a unique and objective truth-value which is independent of whether it can be known by us or proved from our current mathematical theories.

So truth-value realism is clearly a *metaphysical* view. But unlike mathematical platonism, it is not an *ontological* view. For although truth-value realism claims that

mathematical statements have unique and objective truth-values, it is not committed to the distinctively platonistic idea that these truth-values flow from an ontology of independently existing mathematical objects.

What is the relation between these two forms of realism about mathematics? Mathematical platonism clearly motivates truth-value realism by providing an account of how mathematical statements get their truth-values. But further assumptions would be needed for the former view to entail the latter. Even if there are mathematical objects, referential and quantificational indeterminacy may deprive mathematical statements of a unique and objective truth-value (Putnam 1980; Field 1998). Conversely, truth-value realism does not by itself entail mathematical platonism (or even the weaker thesis of anti-nominalism). For we have seen that there are accounts of how mathematical sentences can come to possess unique and objective truth-values which avoid positing a realm of mathematical objects. In fact, many nominalists endorse truth-value realism, at least about more basic branches of mathematics, such as arithmetic.

Nominalists of this type are committed to the slightly odd-sounding view that, although the ordinary mathematical statement

(3) There are primes numbers between 10 and 20.

is true, there are in fact no mathematical objects and thus in particular no numbers. But there is no contradiction here. We must distinguish between the language \mathcal{L}_M in which mathematicians make their claims and the language \mathcal{L}_p in which philosophers make theirs. The statement (3) is made in \mathcal{L}_M . But the nominalist's assertion that (3) is true but that there are no abstract objects is made in \mathcal{L}_p . The nominalist's assertion is thus perfectly coherent provided that (3) is translated non-homophonically from \mathcal{L}_M into \mathcal{L}_p . And when nominalists claim that the truth-values of sentences of \mathcal{L}_M are fixed in a way that doesn't appeal to mathematical objects, it is precisely this sort of non-homophonic translation they have in mind. (The modal structuralism outlined in section 24.3 provides an example.) This shows that the theses of mathematical platonism and anti-nominalism will have their intended effect only if they are expressed in the language \mathcal{L}_p used by us philosophers. If the theses were expressed in the language \mathcal{L}_M used by mathematicians, then nominalists would be able to accept them while still denying that there are mathematical objects.

We have seen that it is possible to accept one form of realism about mathematics without the other. I now survey some of the main philosophical views on the relation between the two forms of realism.

A traditional platonistic view accepts both forms of realism but regards the platonist form as more fundamental. On this view, the objective truth-values of mathematical statements are underpinned and explained by the independent existence of mathematical objects. It is because these objects exist and have their properties independently of us that every meaningful mathematical question has an objective answer. Mathematics is in this respect like the natural sciences, whose statements also have objective

truth-values that are determined by the objects concerned and these objects' perfectly objective properties.

However, this traditional platonist view gives rise to serious epistemological problems. The proposed order of explanation may well be appropriate in the natural sciences, where it makes sense to talk about first identifying certain objects and then examining these objects in order to determine their mind-independent properties. But it is doubtful that it makes sense to talk about identifying an abstract mathematical object independently of determining the truth-values of statements about it. If we have any kind of "access" to such objects at all, this seems to proceed via the truth-values of claims about these objects.¹⁸

These considerations suggest that mathematical platonists should proceed in the opposite direction and regard truth-value realism as more fundamental than mathematical platonism. This reversal of explanatory direction is often associated with Frege and his "context principle," the most famous occurrence of which is found in the *Foundations of Arithmetic*:

How, then, are the numbers to be given to us, if we cannot have any ideas or intuitions of them? Since it is only in the context of a sentence that words have any meaning, our problem becomes this: To define the sense of a sentence in which a number word occurs. (Frege 1953: section 62)¹⁹

The proposal is thus that the question about our "access" to mathematical objects should be transformed to a question about the meaning of complete sentences concerned with these objects.

When implementing this proposal, Frege argued that it is particularly important to explain the meaning of identities flanked by number terms of the form " $\#F$," which abbreviates "the number of F s." He suggests that the meaning of such identities can be explained by what has become known as Hume's Principle:

$$(HP) \quad \#F = \#G \leftrightarrow F \approx G$$

where the right-hand side abbreviates the formalization in pure second-order logic of the claim that the F s and the G s can be one-to-one correlated. This principle, and abstraction principles more generally, have since played a central role in Fregean approaches to mathematics (Wright 1983; Hale and Wright 2001).

¹⁸ A more general epistemological problem was articulated by Benacerraf (1973). How can concrete beings like us gain knowledge of abstract objects with which we cannot be causally influenced? It remains controversial whether this problem of "epistemic access" is genuine; see Burgess and Rosen (1997). The problem described in the text is more specific, as it concerns the explanatory direction associated with traditional platonism.

¹⁹ My translation differs slightly from that of J. L. Austin in that I render Frege's original *Satz* as "sentence" rather than "proposition." This should be uncontroversial, as the context in which a syntactic item such as a word occurs is clearly that of a sentence, not a proposition.

The Fregean tradition regards questions about complete sentences as explanatorily more fundamental than questions about individual singular terms. An important aspect of this orientation is the view encapsulated in Kreisel's famous dictum that "the problem is not the existence of mathematical objects but the objectivity of mathematical statements."²⁰ That is, the sentence-level question of truth-value realism is more fundamental than the question of mathematical platonism, which is concerned with the reference of singular terms. An influential proponent of this orientation is Dummett, who has long urged that the debate about platonism should be replaced by, or at least transformed into, a debate about truth-value realism, because the latter debate is more tractable and of greater importance to philosophy and mathematics than the former (Dummett 1978a: 228–32; 1991: 10–15). Mathematical objects are on this view at best a kind of byproduct of the objectivity of mathematical sentences. Insofar as we can "identify" any mathematical objects at all, this goes via their properties.

This emphasis on the close connection between mathematical objects and mathematical objectivity is shared by some non-Fregean approaches as well. One example is non-eliminative structuralist views, which hold that there are mathematical objects but that these are nothing but positions in abstract mathematical structures. By tying mathematical objects to their structures, such views link the question about the existence of mathematical objects and the question about truth in a structure. Indeed, Shapiro goes as far as to say that eliminative and non-eliminative structuralism "are equivalent" and "[i]n a sense . . . say the same thing, using different primitives" (Shapiro 1997: 96–7).

24.5 DEFENSES OF TRUTH-VALUE REALISM

Truth-value realism claims that every meaningful mathematical sentence has a unique and objective truth-value that is independent of whether it can be known by us or proved from our current mathematical theories. I now discuss two strategies for defending this claim.

One strategy aims to show that in many core branches of mathematics, our theories and conceptions determine a structure that is unique up to isomorphism (in the sense that any two models that satisfy one of these theories and conceptions are isomorphic). A theory or conception with this property is said to be *categorical*. Since the truth-value of a sentence is the same in any two isomorphic models, any categorical theory or conception will thus ensure that every sentence of the relevant language has a unique and objective truth-value.

How can a mathematical structure be determined uniquely up to isomorphism? One popular answer appeals to the categoricity theorems of second-order logic, which say

²⁰ As reported in Dummett (1978b: xxxviii). See also Dummett (1981: 508). The remark of Kreisel's to which Dummett is alluding appears to be Kreisel (1958: 138 fn. 1) (which, if so, is rather less memorable than Dummett's paraphrase).

that categorical characterizations of many of the basic structures of mathematics are available, provided that the second-order quantifiers are given the standard interpretation as ranging over all subsets of the domain (Shapiro 2000a). In particular, on the standard interpretation of second-order logic, categorical characterizations are available of the structures of the natural numbers and the real numbers, as well as of some important initial segments of the iterative hierarchy of sets.

However, the reliance on second-order logic is a cause for concern. The argument assumes that there is a unique standard interpretation of the second-order quantifiers and that we have succeeded in giving our second-order quantifiers this particular interpretation. But what assurance do we have that these assumptions are satisfied? One might respond that the second-order quantifiers are perfectly well-understood and can therefore legitimately be taken at face value. But this response is problematic in the present dialectical situation. For the ordinary language of arithmetic is just as well understood as that of pure second-order logic. So if it is legitimate to take the latter at face value and dismiss worries about non-standard interpretations, presumably the same goes for the former. But then the detour via second-order logic becomes redundant! In short, why should someone who is genuinely worried about the categoricity of arithmetic be any less worried about the interpretation of our second-order quantifiers? It is not a good answer that the second-order quantifiers belong to pure logic whereas the primitive expressions of arithmetic do not. For regardless of the logicity of these expressions, they will have to be interpreted, and mathematical logic shows that a rich variety of non-standard interpretations is available.

For a defense of the categoricity of arithmetic to do the requisite philosophical work, the resources that it employs must be less problematic than those of arithmetic. Several philosophers have recently argued that a schematic form of Peano arithmetic fits the bill. Let me outline the argument as developed by Charles Parsons.²¹ The key observation is that the principle of mathematical induction has a schematic character. When we learn that any property had by 0 and inherited from one natural number to the next is had by any natural number, the notion of property that is used is not tied to what is expressible in a fixed language, nor to any other fixed domain of properties. Rather, whenever we become convinced that a formula $\phi(x)$ can be meaningfully applied to the natural numbers, we accept the corresponding induction axiom:

$$\phi(0) \wedge \forall n(\phi(n) \rightarrow \phi(n+1)) \rightarrow \forall x(\mathbb{N}(x) \rightarrow \phi(x))$$

Assume now that two arithmetically-competent people engage in discourse and each becomes convinced that the other's arithmetical vocabulary is meaningful. The schematic principle of induction then licenses each person to use induction axioms that contain the other's arithmetical vocabulary. The two people can thus go through the steps of the ordinary categoricity argument and prove that their respective "copies" of the natural

²¹ See McGee (1997); Lavine (1994); Parsons (2008: section 49).

numbers are isomorphic.²² Each person can thus convince herself that “her” natural numbers are isomorphic to those of any other person with whom discourse is possible. This restricted form of categoricity appears to be sufficient for most purposes.

The previous paragraphs have focused on the case of arithmetic. What about analysis and set theory? The concerns about the reliance on (standardly interpreted) second-order logic remain largely unchanged. However, it remains unclear whether a schematic approach, with less problematic assumptions, can be made to work beyond arithmetic.

The second strategy for defending truth-value realism aims to show that there are non-arbitrary ways of extending incomplete mathematical theories. Gödel’s incompleteness theorems provide a simple example. Provided that it is consistent, PA neither proves nor refutes the formalization, $\text{Con}(\text{PA})$, of the claim that it is consistent. But as we saw in section 24.3, it is far more plausible to extend PA by adding $\text{Con}(\text{PA})$ than by adding the negation of this formula. In particular, $\text{Con}(\text{PA})$ becomes provable when second-order logic is added to PA. Emboldened by this observation, Gödel sought to generalize. Perhaps all cases of incompleteness can be eliminated by adding expressive and inferential resources of higher and higher order, or (equivalently, Gödel thought) by adding axioms that require the iterative hierarchy of sets to extend higher and higher.²³ Unfortunately, it has turned out that some of the most interesting cases of incompleteness stubbornly resist this attempted elimination. An example is Cantor’s famous Continuum Hypothesis, which is neither provable nor refutable from standard ZFC set theory and remains so even when large cardinal axioms are added. In contemporary set theory the discussion has therefore moved on to more “extrinsic” considerations for and against possible extensions of our mathematical theories. For instance, the evidence that is adduced concerns the extent to which a proposed extension is natural, explanatory, theoretically fruitful, and extends already established patterns.

Perhaps such considerations will one day convince the mathematical community to adopt certain new axioms.²⁴ If so, will this development support truth-value realism by providing an example of how mathematical statements can have objective truth-values despite not being provable from our current theories? The answer will depend on how the convergence on the new axioms is best explained.

One explanation is that the convergence results from an implicit decision by mathematical community to sharpen their concepts and theories in a particular way. A comparison may be instructive. Suppose the chess community was required to explore ways of extending the rules of chess. A convergence on one optimal way of doing so might well result. But the best explanation of this convergence would not be that the chosen additions were true or valid all along, but rather that the resulting extension has

²² Strictly speaking, we also need a schematic form definition by primitive recursion. But this is no more problematic than the schematic form of induction.

²³ See Gödel (1946: 151). For discussion of this point, as well as the material in the rest of this paragraph, see Koellner (2006).

²⁴ This is in no way clear, as there is substantial resistance, as illustrated for instance by Feferman (1999).

attractive properties that moved the community to favor this extension of their game over alternative ones.

Another explanation of a convergence on new axioms sees it as a gradual uncovering of an implicit conception that the mathematical community has of the mathematical structure in question. A comparison may again be instructive. Perhaps the principle of induction was implicit in mathematicians' conception of the natural numbers prior to its first systematic and widespread adoption as an axiom in the seventeenth century. Earlier mathematicians may for instance have been implicitly committed to the view that the natural numbers are exhausted by 0 and numbers that can be reached from 0 by iterated application of the successor function. If so, then the convergence on the principle of induction as an axiom would only support a weak form of truth-value realism because the truth of the principle was already present, if not exactly in mathematicians' explicit theory of the natural numbers, then at least in their implicit conception of this structure.

Thus, even if there were to be a convergence on new axioms for set theory, more work would be needed to show that this phenomenon supports truth-value realism at all (unlike the first alternative explanation outlined above), or that it supports more than a weak form of truth-value realism (unlike the second alternative explanation).

REFERENCES

- Benacerraf, P. (1973). "Mathematical Truth," *Journal of Philosophy* 70: 661–79. Reprinted in Benacerraf and Putnam (1983).
- Benacerraf, P. and H. Putnam (eds.) (1983). *Philosophy of Mathematics: Selected Readings*, 2nd edn. Cambridge: Cambridge University Press.
- Burgess, J. P. (1999). "Review of Stewart Shapiro, *Philosophy of Mathematics: Structure and Ontology*," *Notre Dame Journal of Formal Logic* 40: 283–91.
- Burgess, J. P. and G. Rosen (1997). *A Subject with No Object*. Oxford: Oxford University Press.
- Carnap, R. (1934). *Logische Syntax der Sprache*. Vienna: Springer. Trans. as *The Logical Syntax of Language* (1937). New York: Harcourt, Brace.
- Dummett, M. (1978a). "The Philosophical Basis of Intuitionistic Logic," in M. Dummett, *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press, 215–47. Reprinted in Benacerraf and Putnam (1983).
- Dummett, M. (1978b). *Truth and Other Enigmas*. Cambridge, MA: Harvard University Press.
- Dummett, M. (1981). *Frege: Philosophy of Language*, 2nd edn. Cambridge, MA: Harvard University Press.
- Dummett, M. (1991). *The Logical Basis of Metaphysics*. Cambridge, MA: Harvard University Press.
- Feferman, S. (1999). "Does Mathematics Need New Axioms?" *American Mathematical Monthly* 106: 99–111.
- Field, H. (1984). "Is Mathematical Knowledge Just Logical Knowledge?" *The Philosophical Review* 93: 509–52.
- Field, H. (1991). "Metalogic and Modality," *Philosophical Studies* 62: 1–22.
- Field, H. (1998). "Mathematical Objectivity and Mathematical Objects," in S. Laurence and C. Macdonald (eds.), *Contemporary Readings in the Foundations of Metaphysics*. Oxford: Blackwell, 387–403. Reprinted in Field (2001).

- Field, H. (2001). *Truth and the Absence of Fact*. Oxford: Clarendon Press.
- Frege, G. (1903). *Grundgesetze der Arithmetik*, vol. 2. Jena: Hermann Pohle.
- Frege, G. (1952). *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.
- Frege, G. (1953). *Foundations of Arithmetic*, trans. J. L. Austin. Oxford: Blackwell.
- Frege, G. (1984). *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Blackwell.
- Gödel, K. (1946). "Remarks before the Princeton Bicentennial Conference on Problems in Mathematics," in Gödel (1990), 150–3.
- Gödel, K. (1990). *Collected Works*, vol. 2. Oxford: Oxford University Press.
- Hale, B. and C. Wright (2001). *Reason's Proper Study*. Oxford: Clarendon.
- Hellman, G. (1989). *Mathematics without Numbers*. Oxford: Clarendon.
- Hilbert, D. (1926). "Über das Unendliche," *Mathematische Annalen* 95: 161–90. Trans. as "On the Infinite" in Benacerraf and Putnam (1983).
- Koellner, P. (2006). "The Question of Absolute Undecidability," *Philosophia Mathematica* 14: 153–88.
- Kreisel, G. (1958). "Review of Wittgenstein's Remarks on the Foundations of Mathematics," *British Journal for the Philosophy of Science* 9: 135–58.
- Lavine, S. (1994). *Understanding the Infinite*. Cambridge, MA: Harvard University Press.
- Leitgeb, H. (2009). "On Formal and Informal Provability," in O. Bueno and Ø. Linnebo (eds.), *New Waves in Philosophy of Mathematics*. Basingstoke: Palgrave, 263–99.
- Lewis, D. (1991). *Parts of Classes*. Oxford: Blackwell.
- Maddy, P. (1997). *Naturalism in Mathematics*. Oxford: Clarendon Press.
- McGee, V. (1997). "How We Learn Mathematical Language," *The Philosophical Review* 106: 35–68.
- Parsons, C. (2008). *Mathematical Thought and Its Objects*. Cambridge: Cambridge University Press.
- Pettigrew, R. (2008). "Platonism and Aristotelianism in Mathematics," *Philosophia Mathematica* 16: 310–32.
- Prawitz, D. (2006). "Meaning Approached Via Proofs," *Synthese* 148: 507–24.
- Putnam, H. (1967). "Mathematics without Foundations," *Journal of Philosophy* 64: 5–22. Reprinted in Benacerraf and Putnam (1983).
- Putnam, H. (1980). "Models and Reality," *Journal of Symbolic Logic* 45: 464–82. Reprinted in Benacerraf and Putnam (1983).
- Resnik, M. (1980). *Frege and the Philosophy of Mathematics*. Ithaca: Cornell University Press.
- Rosen, G. (2017). "Abstract Objects," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), <http://plato.stanford.edu/entries/abstract-objects/>
- Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press.
- Shapiro, S. (2000a). *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford: Oxford University Press.
- Shapiro, S. (2000b). *Thinking about Mathematics*. Oxford: Oxford University Press.
- Stanley, J. (2001). "Hermeneutic Fictionalism," *Midwest Studies in Philosophy* 25: 36–71.
- Wright, C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen University Press.
- Yablo, S. (2001). "Go Figure: A Path Through Fictionalism," *Midwest Studies in Philosophy* 25: 72–102.
- Yablo, S. (2005). "The Myth of the Seven," in M. Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press, 88–115.

PART VI

.....

FORMAL THEORIES
AND PARADOX

.....

CHAPTER 25

SEMANTIC PARADOXES

A Psychohistory of Self-Defeat

ROY A. SORENSEN

The road to wisdom?—Well, it's plain
and simple to express:
Err
and err
and err again
but less
and less
and less.
Piet Hein

XENOPHANES (c.570–c.475 BC) may have been the first to utter a semantic paradox:

... and of course the clear and certain truth no man has seen nor will there be anyone who knows about the gods and what I say about all things. For even if, in the best case, one happened to speak just of what has been brought to pass, still he himself would not know. But opinion is allotted to all. ((B34) fragment)

Karl Popper (1962) interprets Xenophanes as an apostle of verisimilitude. We cannot reach the truth but we can get closer and closer.

The simple rationale for inquiry is to end ignorance. Once you know the answer, you stop asking the question. Popper regards “I know” as dogmatic; you are closing your mind to further evidence. He favors endless inquiry.

But this open-mindedness raises Meno's question as to how a *skeptic* can justify inquiry. What is the point of asking if you cannot detect the correctness of an answer?

Karl Popper's answer is that we can measure truthlikeness. When Xenophanes says that rainbows are clouds that rise from earth as vapor, his explanation is closer to the truth than Hesiod's claim that the rainbow is the goddess Iris (a messenger between earth and the gods).

Inquirers approach truth in the way $(n-1)/n$ approaches 1. There can be endless progress even though the truth cannot be reached. Truth is a regulative ideal.

What about this asymptotic doctrine itself? If true, then it is, at most, only close to being true. But if it is close to being true, then it is true. So it must actually be false—not even close to the truth!

Was Xenophanes just being wily? Perhaps he wanted his doctrine to defeat itself. What more poignant way to illustrate the impossibility of full truth? The doctrine of verisimilitude then rises Phoenix-like, from the ashes of self-defeat.

25.1 THE SYSTEMATICITY OF SELF-DEFEAT

Self-defeat is a predictable hazard for truth-theorists. "True" is a *general* term. Each proposition is such that it or its negation is true. Since generalizations about truth are themselves propositions, all theorizing about truth is self-referential.

When a generalization fails to count itself as true, the generalization is self-defeating. Such generalizations might be thought too implausible to attract serious adherents. But it is surprisingly difficult to avoid self-defeat when theorizing about truth. Indeed, several prominent philosophers contend that self-defeat is inevitable.

Patterns of semantic self-defeat are worth studying partly because they shake apart inadequate theories of truth and partly because this disintegration sets off sympathetic vibrations through satellite concepts. The basis for this resonance is clear for concepts that aim at truth (guessing) or which entail truth (knowledge) or which preserve truth (validity) or which employ concepts such as "true of" (sets).

Consider the fallibilist's dictum, "No universal negative proposition is provable." This seems plausible because of the infeasibility of inspecting each F in the domain covered by "No Fs are Gs." However, "No universal negative proposition is provable" is itself a universal negative. So if it were true, then no one would be able to prove it. Anyone who asserts "No universal negative proposition is provable" violates the requirement to assert only what one knows.

Self-applicability is an adequacy condition on the theory of truth: no law of truth yields a falsehood when applied to itself.

25.2 A PHILOSOPHICAL SKILL

Infractions of this meta-law have greatest salience to students of analytic philosophy. One of the founders of this movement, Bertrand Russell, promoted semantic paradoxes from their status as scholastic relics: “A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science” (1905: 484–5).

This elevation affected popular culture. Thanks to the computer revolution, an enormous number of people study logic books. The authors routinely mention the liar paradox “This statement is false.” In the popular imagination, the impossibility of consistently assigning a truth-value to this statement is pictured as a romantic limit on logical thinking. Robots are portrayed as vulnerable to sabotage by the liar paradox because their reasoning is not tempered by intuition. This predisposes many people to accept J. R. Lucas’s (1961) thesis that semantic paradox provides a way of distinguishing man from machine.

The hope of using the liar paradox to separate man from machine dates back to Gordon Dickson’s “The Monkey Wrench,” *Astounding Science Fiction*, August 1951. A man tells the computer, “You must reject the statement I am now making to you, because all the statements I make are incorrect.” The trope is familiar enough to be parodied at the level of mass entertainment. In a 1994 episode of *The Simpsons*, “Itchy & Scratchy Land,” there is a deleted scene in which Lisa tries to cast the logical spell over a mob of attacking robots. (She succeeds only in befuddling her dim-witted father Homer.)

Despite this cultural penetration, most people remain slow to recognize semantic paradox. Learning how to spot semantic paradox resembles learning how to read an analog clock. Mastery makes the process deceptively automatic. You just *see* the correct time by a glance at the clock. Later, as a parent, you are surprised that your child does not experience dissonance when viewing an analog clock in a mirror (Figure 25.1).

The anomaly is invisible to contemporary children who have easy access to *digital* watches. They just lack the practice with analog clocks.

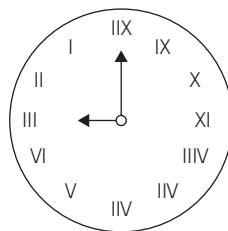


FIGURE 25.1 Mirror image of clock

Automatic processes are associated with innate reflexes. But the Stroop effect demonstrates that automatic processes can be learned. People who can read English have slower reaction times to words that are written in mismatching colors such as **WHITE**. Prisoners who feign illiteracy are betrayed by their slow reaction times to these dissonant stimuli (Lu et al. 2004). To see the color of the word, they must look at it. And the looking triggers reading.

The detection of semantic paradoxes is also automatic. A psychologist who wished to ferret out philosophical literacy could devise a Stroop test that exploits sensitivity to semantic paradox.

25.3 ROBOT HISTORIOGRAPHY

The history of logic confirms that the detection of semantic paradoxes is a skill that was slow to develop and which atrophies when not sustained by at least light practice. The ability to spot semantic paradox marks a philosopher as hailing from one of a small number of eras (rather like the ability to use a slide rule dates engineers).

There may be no indigenous recognition of semantic paradox east of the Euphrates river. Philosophers of India eventually marshal self-refutation arguments (Perrett 1984). However, all recorded instances date from after Aristotle's pupil Alexander invaded India. Since self-refutation arguments do not rely on the propounder's assumptions, they do not face the usual barriers of cultural transmission. A Greek who knew this style of argument could deploy it effectively despite abject ignorance of the Indian philosopher's wider system.

The closest the Chinese come to anticipating the Greeks is with Mo-tzu (490–403 BC). He attempts to refute ethical partialism of Confucius by showing that espousing the doctrine is self-defeating. The setting is a dialogue with Wu-ma-tzu. Wu-ma-tzu declares that he cares for his fellow countrymen more than those of other states, his neighbors more than people from other districts, his family more than his neighbors, and himself more than his parents. In sum, Wu-ma-tzu judges by nearness to himself. Mo-tzu asks Wu-ma-tzu if he will hide his moral views. Wu-ma-tzu says he will tell others about it. Mo-tzu pounces:

In that case, if one man, ten men, the whole world, are persuaded by you, then one man, ten men, the whole world, will wish to kill you to benefit themselves. If one man, ten men, the whole world, are not persuaded by you, then one man, ten men, the whole world, will wish to kill you as a practitioner of dangerous tenets. If whether persuaded or not they wish to kill you, it's a case of "What gets you hanged is your own mouth"; you are the one who by establishing it as a norm gets yourself killed. (Graham 1989: 61–2)

This dialogue is based on a document from about 300 AD reporting remarks from individuals living in the fourth century BC. So if the document is accurate, this would show precocious awareness of the phenomenon of self-defeat.

But it is not an example of *semantic* self-defeat. Mo-tzu has merely shown that Wu-ma-tzu's principle may imply that he ought not to publicize the principle. He ought to keep it secret. The fact that a prediction would be rendered false *if* it were presented a certain way does not show that it actually is false.

Semantic self-defeat would follow if the partialists subscribed to a principle that all moral truths are potentially common knowledge. The moral rules cannot be secrets! Several Western ethicists have argued for this publicity principle. But the partialist should have no more attraction to this adequacy condition than has been displayed by ethical egoists and utilitarians. In the first chapter of *Reasons and Persons*, Derek Parfit characterizes egoism and consequentialism as self-defeating. Each doctrine specifies conditions under which it ought not to be believed. For instance, egoism implies that it ought not to be believed when this belief is not in your self-interest. If a rich altruist will give a poor egoist a fortune for believing that egoism is false, then the poor egoist ought to persuade himself that egoism is false. But this egoist case for defection from egoism is not a refutation of egoism *as theory about what is moral*.

A theory is self-effacing if it advises you not to believe it. If an egoist can live longer and more happily by converting to altruism, then the egoism entails that he ought to convert. That implication would refute egoism only if the egoist says you should believe the theory *under all circumstances*. Egoism is about what is true not about what should be believed.

Egoism is collectively self-defeating in the sense that individual egoists have conflicting goals. When there a pie to divide, each egoist wants the whole pie for himself. In game theory, the prisoners' dilemma shows how egoists will undermine each other when there are public goods.

A common reply is that the prisoners' dilemma merely shows that individual rationality does not imply collective rationality. There is nothing irrational in wrestlers each seeking to prevail over the other. Competition is not contradiction.

Utilitarians make the same moves as the egoists. Like the egoist, the utilitarian is free to distinguish between delivering a correct theory and delivering advice about what to believe.

Picasso characterized art as the lie that tells the truth. If Xenophanes espouses the principle of verisimilitude in this fashion, can he be faulted for straying from the full truth?

25.4 ANCIENT GREEKS

Xenophanes exhibits no fear that his generalizations will indict themselves. Experts in pre-Socratic philosophy, trained in analytic philosophy, infer that Xenophanes must have been exempting his own statement from the scope of the principle. But this is anachronistic. The earliest awareness of self-defeat is in a fragment controversially

attributed to Democritus (Castagnoli 2010) in which he objects “Nothing is true” is self-defeating. The pre-Socratics were trial-and-error pioneers, groping toward the dialectical skills later experts would take for granted.

Even experts lapse when their special (and historically recent) skills failed to be cued into operation. Statisticians commit the same fallacies as amateurs when questions are posed in unfamiliar ways.

Logicians get rusty. Alfred North Whitehead was well versed in the semantic paradoxes when he co-authored *Principia Mathematica* (1910, vol. 1: 63) with Bertrand Russell. But forty years later, Whitehead evinces the serenity of a Professor Emeritus: “There are no whole truths; all truths are half-truths. It is trying to treat them as whole truths that plays the devil” (1953: 14).

Bertrand Russell regarded G. E. Moore as one of the most of truthful men. When Russell asked Moore whether he always told the truth, Moore answered, “No.” Russell conjectured that this was the only lie Moore ever told. Saul Kripke concludes, “Even the subtlest experts may not be able to avoid utterances leading to paradox” (1984: 55). Others counter that Russell must have been making a scholarly joke. But if Russell were focusing on Moore’s *moral* virtue, then the oversight is plausible to experts on expertise.

Historians underestimate the extent to which obviousness is the effect of what K. Anders Ericsson (1991) calls “deliberate practice.” Novices need about 10,000 hours (around ten years) of attentive, self-conscious training (as opposed to mere participation) to develop the automated responses upon which expert intuition is founded. If the expert fails to keep in practice, his performance will revert to the average.

Since historians use the method of empathy and the principle of charity when interpreting past thinkers, they resist attributing obvious errors of omission and commission. But obviousness must be relativized. These relativizations need to be guided by the sort of cognitive psychology summarized by Keith Stanovich’s *The Robot’s Rebellion*. Since proofs build from obvious inferences, the history of logic is a promising beachhead for this automation of intellectual history.

25.5 EPIMENIDES AND THE LIAR

Xenophanes’ chief competitor for being the first victim of semantic paradox is Epimenides—a partly mythical figure of the sixth century BC. Epimenides was vaguely associated with the origin of the liar paradox in the nineteenth century. But he became regarded as a deliberate propounder of the paradox after Bertrand Russell anointed him in “Les Paradoxes de la Logique,” *Revue de Métaphysique et de Morale*, 14 (September 1906), 627–50 (translated as “On ‘Insolubilia’ and Their Solution by Symbolic Logic” in *Essays in Analysis*, 196–204).

Epimenides is the prophet St Paul warns of in a letter addressed to his bishop on the isle of Crete: “One of themselves, even a prophet of their own, said, The Cretans are

always liars, evil beasts, slow bellies. This witness is true" (Titus 1: 12–13.) Paul is quoting from Epimenides' religious poem "Cretica":

They fashioned a tomb for thee, O holy and high one—
The Cretans, always liars, evil beasts, idle bellies!
But thou art not dead: thou livest and abidest forever,
For in thee we live and move and have our being.

St Paul evinces no awareness that it is self-defeating for a *Cretan* to say, "The Cretans always lie." The fact that the Cretans themselves concede they always lie is instead construed as lending extra *credence* (akin to testimony against one's own interest in a court of law).

Nor is there historical evidence that Epimenides was aware of the self-defeat. The context suggests that Epimenides was admonishing his countrymen for impiety. He laments them belittling Zeus (by describing this god as *mortal*).

A few Christians eventually recognize the phenomenon of self-defeat. Augustine taught rhetoric before his conversion to Christianity. Augustine's training facilitated pre-Cartesian meditations on "I do not exist" (Sorensen 2003: ch. 12).

Augustine was writing at the twilight of the Roman Empire. There followed a period of seven hundred years in which Christendom is naïve about the semantic paradoxes. This ignorance persisted despite possession of competent summaries of the liar paradox by Aristotle, Augustine, and Cicero.

There were many thinkers between Epimenides and Paul who were adept at identifying and exposing self-defeating propositions. Consider Plato's critique of Protagoras' slogan "Man is the measure." Plato interprets Protagoras' relativism as entailing subjectivism: "All beliefs are true." He then has his spokesman, Socrates, exploit an internal conflict between beliefs and meta-beliefs:

... there is no one in the world who doesn't believe that in some matters he is wiser than other men; while in other matters, they are wiser than he. In emergencies—if at no other time—you see this belief. When they are in distress, on the battlefield, or in sickness or in a storm at sea, all men turn to their leaders in each sphere as to God, and look to them for salvation because they are superior in precisely this one thing—knowledge. And wherever human life and work goes on you find everywhere men seeking teachers and masters, for themselves and for the other living creatures and for the direction of all human works. (*Theaetetus* 170 b–c)

Consider someone who believes that at least one of his own beliefs is false. Protagoras' principle that all beliefs are true implies that "At least one of my beliefs is not true" is true for this modest individual. Thus "All beliefs are true for the believer" when applied to the modest believer yields "Not all beliefs are true for the believer." Since Protagorean relativism implies its own negation, Protagorean relativism is false.

Socrates' objection to Protagoras is a premonition of D. C. Mackinson's (1965) preface paradox: a modest author opens with an apology for the errors that are bound to be in

the text. This acknowledgment of fallibility is wise. The surprise is that this self-insight precludes the possibility of all of the author's beliefs being true. If the belief expressed in the preface is true, then one of the beliefs in the text is false. If all the beliefs in the text are true, then the belief in the preface is false. Either way, the author has a false belief.

Although the preface paradox damages Protagoras' relativism, it also challenges Plato's methodological assumption that rationality implies consistency (which naturally motivates the expunging of inconsistencies through dialectical examination). If it is impossible for all of the authors' beliefs to be true, then they are jointly inconsistent. Yet the belief expressed in the preface is rational. Plato evinces no awareness of this self-inflicted wound.

Greek playwrights exploit the irony of self-defeat. The popularity of plays makes it likely that there was widespread awareness of semantic paradox in Greek culture. That would explain why the tale about Epimenides reached the ears of St Paul. Appreciation of the self-defeating aspect of a Cretan warning "All Cretans are liars" was lost in the transmission.

25.6 MATURATION OF THE LIAR PARADOX

Self-defeat is not sufficient for contradiction. Obviously, some Cretan has at some time asserted something that was not a lie. Epimenides' "The Cretans always lie" is merely false.

However, a philosopher from the fourth century BC, Eubulides, appears to have noticed a logical peculiarity: "The Cretans always lie" *entails* that some Cretan is not a liar. What if there had been no honest Cretans? What if Epimenides had been the only Cretan and "The Cretans always lie" was his sole remark? Then the statement would be false if true and true if false!

In principle, we could instead reject the assumption that a lone Cretan is possible. Putting the paradox in the first person can close this loophole: "I am lying."

But there is still a hitch. Some lies accidentally turn out true. Formulating the paradox directly in terms of falsehood can close this further loophole: "I am saying something false."

As a further safeguard, we can interpret the "I" as referring to the sentence itself (so as to not assume the existence of any speakers). This yields the classic

L: This statement is false.

Assume L is true. Then L entails it is false. Therefore, L is false. But if L is false, then it is correctly reporting its falsehood. When a statement says only what corresponds to the facts, then it is true. Conclusion: L is true if false and false if true.

Phew! All of that streamlining was long, intricate work. The simplicity of the end product disguises the many stages of development behind a fluent forked tongue.

Further eloquence proved possible. Later philosophers attacked the liar by insisting the statement was neither true nor false. Their truth-value gap solution inspires a leaner variation of the paradox: “This statement is not true.” This does not require the law of bivalence (that every proposition has exactly one truth-value, true or false).

“Revenge liars” incorporate a proposed solution method into a fresh attack. Medieval philosophers were aware of how new liar paradoxes could be created from the toolbox used to “solve” the old liar paradox (Bradwardine 2010: 25). However, only Roger Swyneshed (writing in the early 1330s) seriously entertained the possibility of truth-value gaps (Spade and Read 2013: section 3.2). He has separate discussions of “This proposition is false” and “This proposition is not true.”

Another developmental milestone was to free the paradox from reliance on negation. In 1942 Haskell Curry proposed a material conditional:

C: If this sentence is true, then all sentences are true.

If the antecedent of C is false, then C is true by virtue of being a material conditional with a true antecedent. If the antecedent of C true, then all sentences are true by virtue of *modus ponens*.

Pseudo-Scotus (John of Cornwall) mentions analogues of the Curry paradox framed in terms of validity (Read 1979). Some variations can be freed from “This argument is valid, therefore, something is distinct from itself.” It is impossible for the premise to be true and the conclusion false. So the argument is valid and should therefore be sound. Yet the conclusion is necessarily false.

The medievals would have seen no significant difference between paradoxes that turn on the consequence relation and those that turn on conditionals. They follow Aristotle in neglecting the distinction. This complements their neglect of the distinction between “false” and “not true.” These distinctions only acquired importance with the rise of deviant logics—which echoed the development of non-Euclidean geometries.

The Curry paradox makes little difference to classical logic. However, it is a counterexample to solutions proposed by deviant logicians. For they blame the liar paradox on classical negation or on laws of logic that entitle inference from falsehood.

25.7 A COMPARISON WITH CONTRADICTIONS

The liar appears to contradict itself. But intuitive self-contradiction can be stimulated by sentences that are contingent. For instance, Moore’s problem is to explain why “It is raining but I do not believe it” appears self-contradictory even though it is consistent.

In classical logic, a contradiction entails everything. If the contradiction is true, then all hell breaks loose. If false, hell stays where it is. So we have the stable option of

regarding a contradiction as false. The liar deprives us of this equilibrium. For if L is false, it seems to be synonymous with the true statement:

M: L is false.

The medieval logician Jean Buridan's solution was to challenge this synonymy. His differentiator is the narcissistic principle that every statement implies its own truth—a principle that continues to attract adherents such as Eugene Mills (1998). Whereas M ascribes truth to M, L ascribes truth to L. Since L (implicitly) ascribes truth to itself while explicitly ascribing falsehood to itself, L is a contradiction. Thus we can consistently regard L as false and M as true.

Contradictions are necessarily false. So if the liar contradicts itself, then so should the possible liar:

PL: PL is possibly false.

The dilemma is as follows: if PL is true, then we can consistently suppose PL is false. But if PL is false, then it is necessarily true. So it would be both true and false.

Jean Buridan grabs the second horn of the dilemma. According to Buridan, proposition P implicitly ascribes truth to itself. So it is just a contradiction.

25.7.1 Truth-teller

If Buridan is correct there is an illuminating connection between the liar paradox and the truth-teller "This statement is true." The truth-teller explicitly does what normal propositions only do implicitly: self-attribute truth.

Those fond of symmetry might expect Buridan to conclude that the truth-teller is true. Just as the liar is a subtle contradiction, the truth-teller is a subtle tautology. The truth-teller's empty nature conforms to the stereotype of a tautology as an empty remark.

Historically, however, all of those attracted to the principle that all statements self-attribute truth have instead denied the truth of the truth-teller. Some of Buridan's contemporaries tried to force the truth-teller to say more (Read 2006: 314–15). They resemble prosecutors who claim that silence is a gesture of assent. The silent defendant is interpreted as capitulating to a charge when he declines to contest it. Under this policy of forfeiture by silence, any proposition that gives us no more reason to say p rather than q winds up saying both. Since "This is true" gives us as much reason to describe it as false as to describe it as true, it describes itself both ways and so is a contradiction.

Buridan does not endorse the verdict of this kangaroo court. He prefers no explanation over a bad explanation. The tight-lipped Buridan provides no support for his principle the each declarative statement self-attributes truth.

The most common thirteenth-century solution to the truth-teller was to claim that it says nothing. But some statements seem to say even less than the truth-teller. After all, the truth-teller implies its existential generalization, “Some statement is true.” Since the reverse implication does not hold, the truth-teller has more content than its existential generalization and so must say something.

The truth-teller is also stronger than the possible truth-teller:

PT: It is possible that PT is true.

PT is in turn stronger than the existential generalization “Some statement is possibly true.” This de re statement is stronger than the de dicto “It is possible that there is a true statement.”

One might consider an endless sequence of ever-weaker truth-tellers—perhaps by iterating the possibility operator. Where there are differences in meaning there is meaningfulness. So all the statements would be rendered meaningful!

One might halt this sequence of vaguer and vaguer statements by adopting the modal system S4 (or the stronger S5). S4 has the collapse theorem $\Box\Box p \rightarrow \Box p$. That might yield a weakest truth-teller. But its slightly stronger variants would survive.

This principle is plausible for metaphysical possibility. But it is less plausible that *epistemic* possibility collapses. Under the epistemic reading, $\Box p$ means “p is consistent with all that is known.” Under cross-examination, a witness answers “Maybe and maybe not” to “Might your evidence be compatible with another possibility?” The witness’s answer does not entail an affirmative answer to the lawyer’s question.

If there were a default-truth value, then the truth-teller could receive that truth-value. But should true or false be the default? On the one hand, N. L. Wilson’s principle of charity maximizes the attribution of “True” to assertions. This suggests that the default should be truth. On the other hand, the attribution of falsehood has the advantage of being less committal. Calling a proposition true commits you to the truth of all of its consequences. This suggests truth is an achievement rather than the “rest state” of a proposition.

25.8 THE NO-NO PARADOX

These indecisive musings are pre-empted by an analogue of the truth-teller called the no-no paradox (Sorensen 2001: ch. 13). An early specimen is preserved as a handwritten note in the Bertrand Russell archives:

G. G. Berry, the writer of the following letter, was a man of very considerable ability in mathematical logic. He was employed in a rather humble capacity in The Bodleian, his subject being one that the University of Oxford ignored. The first time he came to see me at Bagley Wood he was bearing, as if it were a visiting card,

a piece of paper on which I perceived the words: "The statement on the other side of this paper is false." I turned it over & found the words: "The statement on the other side of this paper is false". We then proceeded to polite conversation. (Quoted by Garciadiego 1992: 166)

Russell intended that one of the sentences attribute *truth* to the other while the other attributes falsehood to the other. He repeats the slip of the pen in the 1951 edition of his autobiography.

Readers of the autobiography alerted Russell to the slip. He published a correction in a letter to *The Observer* "False or True" (12 March 1967: 33). This correction was incorporated into the second printing of his 1967 edition. However, the slip reappears in the 1971 paperback (and therefore the slip of the pen gets the last word because Russell died in 1970).

The reverse slip also occurs. The first edition of R. M. Sainsbury's *Paradoxes* correctly quotes Jean Buridan's Eighth Sophism:

Socrates in Troy says, "What Plato is now saying in Athens is false." At the same time, Plato in Athens says, "What Socrates is now saying in Troy is false." (1988: 145)

However, in Sainsbury's second edition (1995: 145), the last occurrence of "false" is miscorrected to "true." (His third edition corrects the miscorrection. I look forward to the fourth edition.)

Russell did not notice that his slip of the pen is philosophically interesting. As with the truth-teller, there are two consistent assignments of truth-values. However, the assignments are *asymmetrical*. The symmetry between the sentences makes either assignment arbitrary. Privileging one truth-value does not help because the only consistent options are between asymmetric truth-values.

The no-no paradox illustrates a second way automaticity affects the history of logic. It is commonly produced as a slip of the tongue when one is intending to present a looped liar:

L1: Sentence L2 is false.

L2: Sentence L1 is true.

The L2 response is conversationally rare. The more common response is a symmetrical "He said, she said" quarrel in which each party accuses the other of falsehood. Since automatic behavior errs in the direction of a higher base rate, "false" gets substituted for "true" to yield the no-no paradox.

Sainsbury's reverse slip stems from his erudition rather than the ordinary force of habit that repeatedly trips Russell. Sainsbury is akin to a speaker who over-corrects grammar. This hyper-urbanism is abetted by the high base rate of the looped liar in literature on the liar paradox. At the time Sainsbury wrote, there was far more commentary on the looped liar than on the rarely mentioned no-no paradox. The only references to

the no-no paradox between Buridan and Sainsbury were by Arthur Prior (1976: 131) and Lawrence Goldstein (1992).

25.9 THE ROLE OF PARITY

The intimate connection between indirect liar paradoxes and the no-no paradox is underscored by an elongated, Janus-faced sequence:

1. Statement 2 is false.
2. Statement 3 is false.
- \underline{n} -1. Statement \underline{n} is false.
- \underline{n} . Statement 1 is false.

When \underline{n} is an even number, the result is a no-no paradox. When \underline{n} is an odd number, there is a liar paradox.

The parity of the sequence takes on the same surprising significance as the parity of interlocking gears (Figure 25.2)

If the number of gears is odd, they gridlock (because each gear turns its neighbor in the opposite direction). If the number of gears in a circuit is even, then all of them can turn. (See Figure 25.3.)

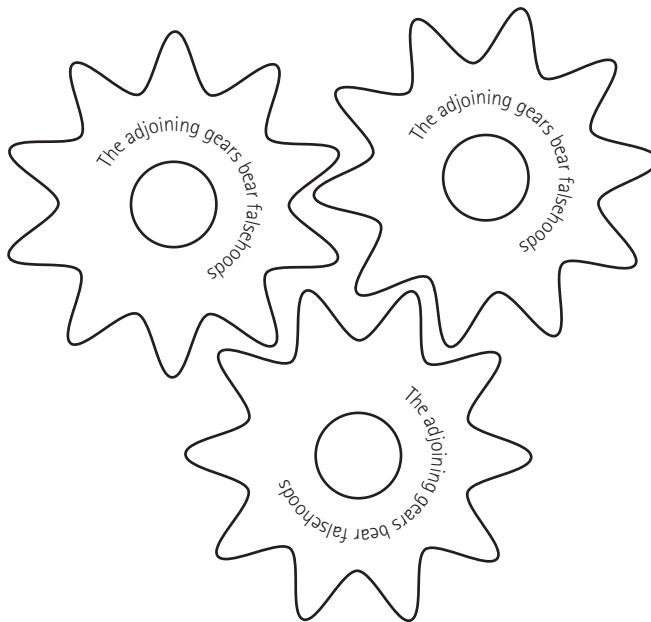


FIGURE 25.2 Three-gear liar paradox

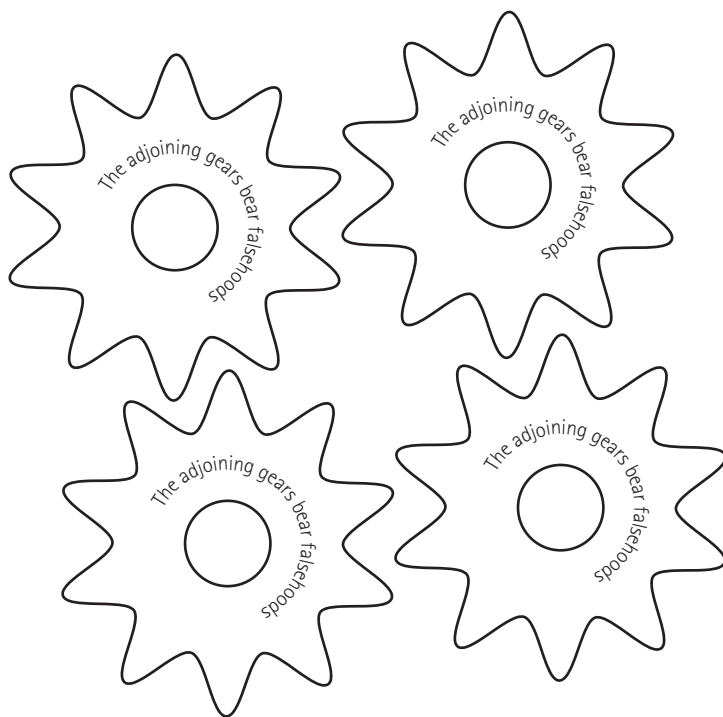


FIGURE 25.3 Four-gear no-no paradox

This isomorphism suggests a way of mechanizing the relationship between the looped liar and the no-no paradox. Each gear bears the inscription “The adjoining gears bear falsehoods” (understood as its *immediately* contiguous gears—the ones they touch). Let clockwise rotation signify truth and counterclockwise rotation signify falsity. Circuits with an even number of gears represent the no-no paradox. They are able to turn in either direction (yielding opposite but equally consistent assignments of truth-values). Circuits with an odd number of gears symbolize the looped liar by gridlock.

25.10 SUBSTITUTING INFINITY FOR CIRCULARITY

In “Paradox without Self-Reference” Stephen Yablo uses an infinite sequence of *universal* generalizations to generate a liar paradox:

1. All of the subsequent sentences are false.
2. All of the subsequent sentences are false.
3. All of the subsequent sentences are false.

Just as in the circular liar L, there is no consistent assignment of truth-values to the sentences in this sequence. If sentence 1 is true, then sentence 2 correctly describes all of its successors as being false. So 2 would be both true and false. If sentence 1 is false, then there must be a true successor, sentence \underline{n} . But if sentence \underline{n} is true, then sentence $\underline{n} + 1$ correctly describes all of its successors as being false. So $\underline{n} + 1$ would be both true and false.

The technique of substituting infinity for circularity was pioneered by Hans Herzberger's infinite truth-teller (1970: 150).

1. The next sentence is true.
2. The next sentence is true.
3. The next sentence is true.

Weakening each sentence to "At least one of the subsequent sentences is true" can also make Herzberger's point or strengthening each sentence to "All the subsequent sentence are true."

An infinite liar can be reached by a slip of the tongue in which "true" is switched with "false" in either of these generalized versions. Yablo's version is a single word substitution from "All the subsequent sentences are true" to "All the subsequent sentences are false."

Stephen Yablo (1985) cites the article in which Herzberger presents the infinite truth-teller and acknowledges his help in discussion. Herzberger's truth-teller sat around like a loaded gun, waiting to be tweaked—or mis-transcribed—into a more powerful paradox.

Not all of the loaded guns fired. Yablo's review of ungrounded sequences contains the following:

- Y1: There is a subsequent pair of adjacent sentences that are equivalent.
 Y2: There is a subsequent pair of adjacent sentences that are equivalent.
 Y3: There is a subsequent pair of adjacent sentences that are equivalent.

Yablo notes that the sequence has a single consistent assignment of truth-values, namely, evaluating each as true. Years later, Roy Cook (2014: 15–16) substituted "false" for "equivalent" to obtain a paradoxical sequence. Bang!

Recognizing that one has inadvertently produced something significant is not robotic. Yablo initially thought little of the paradox because it was not a counterexample to any of the recent theories of truth. Later, he viewed the theories as lucky. The infinite liar conforms to the theories without confirming them.

25.11 CONTINGENT LIARS

Picture an infinite history of mildly revisionary thinkers. Each thinks, "At least one falsehood has been thought in a previous generation." As it turns out, each generation of

these circumspect individuals has said only that. Thus each of their remarks is enmeshed in a liar paradox.

Yet each generation has been the model of caution. Each avoids self-reference by restricting their claim to previous generations.

Contingent liars show that the paradoxicality is not an *intrinsic* property of sentences. If I say “The next sentence you say is true” and you say “Leaves fall when fall leaves,” then both sentences are unproblematically true. But if you instead say “What you just said is false,” then you embroil my remark in a liar paradox.

The extrinsic character of the paradox shows that careful thinking is compatible with paradox. We cannot reasonably abstain from assertions whose truth-values depend on the statements made by other people. We must risk semantic paradox.

25.12 VARIATIONS OF MEANINGLESSNESS

According to Gilbert Ryle, “This very statement is false” is meaningless because it is impossible to completely specify what the demonstrative “this statement” refers to. Substituting the sentence itself just produces another ungrounded demonstrative:

“This statement is false” is false.

No amount of substitution will get rid of “this”!

Ryle’s approach extends to the truth-teller paradox and the no-no. However, Willard Quine has a reformulation of the liar that avoids demonstratives:

The problem is to devise a sentence that says of itself that it is false without venturing outside the timeless domain of pure grammar and logic. Here is a solution:

(3) “Does not yield a truth when appended to its own quotation” does not yield a truth when appended to its own quotation. (Quine 1987: 148)

The eleven-word quotation is a noun rather than a demonstrative. So the reformulated liar is put out of reach of Ryle’s objection.

Incidentally, this feat of self-reproduction led computer programmers to use “quine” to denote any program that generates a copy of its own source text as its complete output. Here is one that works in LISP:

```
((lambda (x)
  (list x (list (quote quote) x)))
 (quote
  (lambda (x)
    (list x (list (quote quote) x))))))
```

25.13 REJECTING UNIVERSALITY

Quine's ingenuity illustrates the richness of natural language. Logicians have trouble persuading linguists that the liar sentence is ungrammatical. The liar sentence passes every standard test of grammaticality—and, thanks to the ingenuity of grammarians, even some of the tests specially designed to flunk the liar.

Intuitions about how language works are as fallible as intuitions about digestion. Although all healthy people are linguistically competent, special training is needed for meta-linguistic competence.

There is a metaphysical reason to expect that linguists will always regard *contingent* liar sentences as grammatical. Linguists commonly conceive of grammar as formalizing a speaker's competence. This makes grammaticality an intrinsic property of the speaker. All the linguist needs to ascertain is whether a sentence is grammatical is to study the speaker's mind.

The last reason to think that the liar will never be judged ungrammatical is that natural languages permit neologisms. If *L* is a natural language, then one can add a word and still have a natural language. Natural languages are universal; they let you talk about anything.

The problem is not that natural languages are unclear about which sentences are grammatical. The problem is that they clearly do permit liar sentences.

So logicians contemplate switching to a less expressive language. The liar paradox cannot be expressed in restricted languages such as chess notation. This suggests a strategy. Enrich an artificial language so that one can express everything worth saying but do not make it rich enough to formulate semantic paradoxes.

Alfred Tarski showed how one could supplement the power of a semantically restricted language so that it approximates the expressivity of a natural language. He claims that this language is sufficient for all scientific purposes.

However, one purpose of scientific language is to provide a medium for creative thinking and understanding; the scientist thinks by talking to himself. Some of this cognitive auto-stimulation is with supplemental notation such as scientific notation. But most of it is with ordinary language.

Is thinking "out loud" just a mental crutch? Is it just a familiar shorthand notation that could be replaced with Tarskian longhand?

Doubtful. Reflection on the liar paradox has played a key role in Kurt Godel's incompleteness result and Alan Turing's Halting Theorem. Both theorems mimic liar sentences without igniting pathological behavior. To approximate the effects of the liar the censor must understand both sides of the approximation.

The issues raised by Tarski's artificial language are prefigured in debates on censorship. One of the censor's dilemmas is that he needs access to the forbidden material to censor it. But if he is not corrupted by the doubtful material why does he think others would be corrupted by it? The difficulty also occurs at the level of dissemination.

The governed must be given fair notice of what is illegal. This means an obscenity law must publicly specify the banned obscenities. Some American publishers have refused to print the prohibition. They reason that publishing the law would violate that very law.

Tarski must rely on natural language to formulate and disseminate his artificial language. If natural language is good enough to erect this indirect means of expression, why isn't it good enough for us to directly think with it? Tarski's artificial language differs from decimal notation which is merely an adjunct to natural language that enhances its efficiency for a special purpose. Tarski's language *reduces* what can be said by depriving us of a central concept—unrelativized truth.

Is Tarski is proposing an analogue of George Orwell's *Newspeak*? Defenders of a stratified approach to truth regard this question as alarmist. Michael Glanzberg (2015) compares truth with proof. Kurt Godel showed that we cannot formalize a single notion of proof for (e.g.) Arithmetic. In practice, our formalizations of proof form a hierarchy. This has not proved a handicap. Initial, heavy-handed, restrictions are typically replaced with lighter restrictions.

Even so, admiration for Tarski's artificial language is predicated on our fluency in the natural language it approximates. He displays recognition of this limit. Instead of claiming to solve the liar paradox, he offers a work-around.

Tarski thinks that the liar paradox is impossible to solve. A *solution* would preserve the universality of natural language. But this feature allows us to formulate the liar sentence. So natural language is inconsistent.

A characteristic feature of colloquial language . . . is its universality . . . But it is presumably just this universality of everyday language that is the primary source of all the semantical antinomies, like the antinomies of the liar or of heterological words. These antinomies seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent. (Tarski 1944: 347–8)

Philosophers normally think of language as a means by which we make assertions—not as making any assertions on its own. This silence precludes inconsistency.

However, Tarski appears to regard analytically true statements, tautologies, as assertions of a language. For instance, English allowed Kurt Grelling to define “heterological” to cover those terms that do not describe themselves such as *monosyllabic*, *symmetric*, and *abbreviated*. The neologism is intuitive and easy to teach, especially when paired with a partner neologism: a term is “autological” if and only if it describes itself. “Pentasyllabic” is autological because “pentasyllabic” is pentasyllabic. Grelling's riddle is, “Is ‘heterological’ heterological?” Listing it as a heterological term makes it self-descriptive and so provides decisive grounds for reclassifying it as autological. Yet classifying it as autological makes “heterological” not describe itself and so triggers the opposite re-classification. Tarski's idea is that “‘Heterological’ is heterological” is both analytically true and analytically false. Thus English is inconsistent.

To interdict the contradiction, Tarski replaces the single, universal “true” with a mass of predicates that resemble cellular slime molds. The slime mold’s constituents spend most of their lives as individual unicellular protists. But upon release of a signal, they assemble into a cluster that acts as one organism. Tarski’s “true-in-language L” predicates form an infinite hierarchy. Individually, they behave much like “true” with respect to disquotation and its service as a sentence variable. Collectively, they approximate the universality of “true.” Unlike the ordinary “true,” the individual truth predicates are relativized to a language. They are meta-linguistic predicates. One of their functions is to remove quotation marks from statements made in lower level languages. Another function is to serve in generalizations about these lower level statements. For instance, statements in the object language obey the laws of logic.

Well, a bit more carefully, they obey laws of logic that are restricted to the object language. To approximate the universality of these laws, we need an infinite hierarchy of these laws. “The” law of excluded middle becomes approximated as an infinite sequence: each sentence of the form “P or not P” is true in the object language, each sentence of the form “P or not P” is true in the meta-language, each sentence of the form “P or not P” is true in the meta-meta-language, and so on.

The liar sentence, at least in the guises Tarski was familiar with, cannot be formulated in this grammar. Nor can the heterological paradox.

The explanation of why they cannot does not take place within these artificial languages. This restriction on the narrative is a telling symptom of the expressive poverty of non-universal languages; one cannot tell the tale of why the restrictions are needed!

Tarski does not disparage natural language. He just has a tragic vision of it. According to Tarski, self-defeat is inevitable when formulating the laws of truth. For the universality of the ordinary notion of truth guarantees that the laws will be powerful enough to express the liar sentence. Classical logic will then take over to deliver a contradiction.

According to Hans Herzberger (1967), this attribution of self-defeat is itself defeating. A language is inconsistent only if its set of tautologies is inconsistent. But that means that at least one of these tautologies is not true. All tautologies are true. So to attribute inconsistency to a language is to claim that there is a sentence that is both true and not true. Thus Tarski’s explanation of why inconsistency is inevitable is itself inconsistent.

25.14 IDEOLOGICAL DEVIANT LOGIC

If Tarski’s pessimistic conclusion that the liar paradox is insoluble is incoherent, then we are justified in seeking a solution rather than merely managing the problem. The solution would require revising our assumptions about how language or logic operates. Tarski dismisses the revision of logic as being too radical to contemplate.

Revisionists generally respond by attacking the credentials of classical logic. For instance, some allege that it was designed for mathematics and so does not extend more broadly.

An unexplored alternative would be to portray Tarski as himself a revisionist. In *one* sense, Tarski is condemning deviant logic. Whereas a supplemental logic (such as modal logic) adds new theorems, a deviant logic subtracts theorems. Susan Haack interprets this subtraction as a denial of classical theorems. But this overlooks subtraction by expurgation—denying the very vocabulary needed to express a classical theorem. Tarski could be a deviant logician in this second sense.

Ditto for Peter Geach. He rejects the concept of absolute identity. Geach believes that identity must always be relativized to a sortal. Consequently, he regards Leibniz's law of identity as an ill-formed formula. This makes Geach a special kind of deviant logician—an ideological deviant.

Quine talks of "deactivating" absolute truth. He insists that "true" be relativized to a language just as Geach insists that "is" be relativized to a sortal. "True" is arguably just as much a logical word as "is." The laws of logic are formulated with "true": "Every proposition is true or false," "Every formula of the form ' p or not p ' is true," and so on. "Valid" is also defined with "true." Thus deviant logicians might be tempted to "out" Quine as a closet deviant—especially when he admonishes them for being deviants.

Ideological deviance avoids the deviant logician's dilemma. According to Quine (1970: 80–3) the deviant logician can disagree with classical logic only by denying one of its theorems. To avoid equivocation, the denial must be in the same vocabulary. But since the meaning of the vocabulary is dictated by the truth-table, their agreement means assent to classical theorems.

Quine eventually concedes that he has trouble impaling the intuitionist on the dilemma. The intuitionist does not deny the law of double negation. He merely declines to assert it.

Arguably, Geach is more revolutionary than the intuitionist. Geach rejects " (x) ($x = x$)" as ill-formed. He is a cousin of A. J. Ayer's non-cognitivist. Whereas the atheist denies "God exists" and the agnostic says he does not know, the non-cognitivist says that "God exists" is meaningless. Non-cognitivists share fewer presuppositions with theists than atheists. If Tarski's rejection of absolute truth is comparable to Geach's rejection of absolute identity, then each rejects classical logic in a fashion that is more radical than the deviant logics Tarski considers beyond the pale. A deviant logician who accepts absolute truth could portray himself as more moderate than Tarski—the closet *ideological* deviant.

25.15 THREE-VALUE LOGIC

The paradigm case of a deviant logic is three-value logic. Jan Łukasiewicz introduced the third truth-value, neuter, as a solution to logical fatalism. This intermediate truth-value was soon applied to the liar paradox.

The postulation of a third truth-value differs from a truth-value gap. The third value is intended to preserve the truth-functionality of logic.

Three-value logic also differs from Xenophanes' idea that statements vary in how close they come to the truth. "Pi equals 3" is closer to the truth than "Pi equals 4" but both are false. If two statements have the neuter truth-value then neither is false.

In principle, one could hybridize verisimilitude with three-valued logic to get locutions such as "close to being neuter." However, the hybrid would not have the motivation of solving Meno's paradox. One of the disadvantages of postulating a third truth-value is the dearth of intuitions to guide the development of its logic.

This absence of constraint can also be an advantage. Assigning truth-values on the basis of mere consistency has an artificial feel. Consider the following logic exercise (Honsberger 1991: 235):

In a certain multiple-choice test, one of the questions was illegible, but the choice of answers, given below, was clearly printed. What is the right answer?

- (a) All of the below.
- (b) None of the below.
- (c) All of the above.
- (d) One of the above.
- (e) None of the above.
- (f) None of the above.

The official solution is (e) because it is the sole consistent answer. But many insist that this answer works only because the illegible test connects with reality.

This worry about groundedness is not stimulated by the novel truth-value of neuter. Intuitions about neuter are weak. So the logician has little friction with common sense. He has a blank canvas for the construction of the new truth-tables.

Given that the negation of a statement with a neuter truth-value is itself neuter, we can view the liar paradox as the negation of the truth-teller. And given that the truth-teller has this neuter value, then so will the liar.

For some, this is an inspired compromise. Instead of letting two-valued logic force them into viewing reality as black and white, they let in some gray!

25.16 VERISIMILITUDE'S REVENGE

Theorizing often takes on an economic aspect. The theorist trade-offs between principles and intuitions, seeking a good fit. This is a decision-theoretic outlook; the choices concern states of nature.

When theorizing about semantic paradoxes, this framework becomes *game-theoretic*. For the paradoxes adapt to one's theorizing—borrowing aspects of the solution to reformulate the problem:

N: This statement is not neuter.

If N is neuter, then it is true and so not neuter. If it is not neuter, then it is true. If we add a fourth truth-value, the paradox-monger will incorporate this novel truth-value into another strengthened liar.

Proponents of truth-value gaps have suggested that we instead assign no truth-value to the liar sentence. But this invites a liar that incorporates the notion of a truth-value gap: "This statement is not true."

Parallel revenge occurs if we say that the liar is meaningless or violates a given semantic rule:

M: This statement is meaningless.

The revenge phenomenon carries over to the interrogative. When asked the trick question, "Will you answer no to this question?" you can answer, "There is no yes-or-no answer." The riddler follows-up: "Is there a yes-no answer to this question?" It is self-defeating to answer no. So the answer must be yes. But this response is ungrounded. There is no way to specify what the answer is.

The riddle can be rephrased with terminology from the logic of questions: a question presupposes P if and only if P is a necessary condition for there to be a true answer to the question. The riddle is, "Does this question have a false presupposition?"

According to Piet Hein, "Problems worthy of attack prove their worth by fighting back." New solutions to the liar are published despite common knowledge that they are vulnerable to the revenge problem. By being a problem for everyone, it becomes less of an objection to anyone in particular. If the revenge problem is one day solved, future historians will marvel at our willingness to publish stillborn theories. All the interest appears to be in the autopsy!

Xenophanes might have been at first staggered by the self-defeat of "Statements can only come close to the truth." However, after contemplating the revenge phenomenon, he might console himself: at least his principle correctly implies its own falsehood. Other generalizations are further from the truth. They lack the self-corrective aspect of a self-defeating generalization. Thus the principle of verisimilitude foretells and diagnoses its own deviation from truth. It is a telling instance of itself.

But wait! The doctrine of verisimilitude now appears to exploit its own self-defeat. For if the principle is close to the truth by virtue of its modest form of self-defeat, then the principle accurately describes itself!

Since the verisimilitude truth-teller implies falsehood in the same breath as it ascribes truth-likeness, it is as much a liar as a truth-teller:

V. This statement is truth-like.

If V is truth-like, then it accurately describes itself and so is true. But if it is true, then it is not merely close to true. So it is false. But how false? Not too false because it accurately describes itself.

25.17 SELF-EFFACING THEORIES

Under what conditions is a theory of *truth* self-defeating? When is a theory of truth, to borrow Parfit's phrase, "absurd on its own terms?" One answer: when the theory implies is its own falsehood.

But what if a theory merely implies its inaccessibility? There does not seem to be any guarantee that everybody should be able to recognize the correct theory. Jean Buridan noticed that some self-referential sentences are selectively incredible:

Y: You do not believe this very sentence.

Although *you* cannot believe Y, those observing you can believe Y. The "cannot" seems analytic here. If so it is the kind of sentence Tarski thinks of language as asserting.

There are analogues of Y that expand the list of those who cannot believe it. The banned people cannot believe the sentence but the rest of us can—partly in virtue of them being banned! By degrees, we reach a universally inaccessible sentence:

A: No one believes this sentence.

Sentence A universalizes the paradox. Our inability to believe A makes it true. Yet recognition of a sentence's truth constitutes belief.

Some solutions create two classes of assertible sentence. For instance, Tim Maudlin (2004: 51) says we can assert sentences whose truth is grounded in the world. We can also assert that they are true or false. But we are not permitted to assert the truth of ungrounded sentences. We are permitted to assert that the liar sentence is untrue. We are not permitted to assert that it is true that "The liar sentence is not true."

Maudlin's abstemiousness is novel. Philosophers are already familiar with situations in which we are inclined to assert that a sentence is true even though we are not inclined to assert the sentence itself. When a smart student asks what he missed in his projective geometry class, a dull friend with a good memory reports Desargues's theorem: "In a projective space, two triangles are in perspective axially if and only if they are in perspective centrally." The dull informant does not grasp the distinction between the two ways of being in perspective, so he does not believe what he is asserting. But he can still be confident it is a true sentence. Maudlin is interested in the converse case. He is claiming there are occasions on which one is entitled to assert the sentence but not entitled to assert it is true. The speaker fully understands the sentence and is fully warranted. Yet he will not vouch for its truth.

Maudlin's theory is designed to imply that the liar sentence is ungrounded. Accordingly, Maudlin asserts that the liar is not true. But if you ask Maudlin how he knows the liar sentence is untrue, he denies that he knows any such thing. Knowledge implies truth and under Maudlin's theory it is not true that the liar sentence is untrue.

Maudlin cannot accept standard rules of assertion such as “Assert only what you know” and “Assert only what is true.” The same goes for standard theories of propositional attitudes. For they accept principles of the form “To have attitude A to p is to have attitude A to ‘ p ’ is true.”

25.18 ACCEPTANCE AND VERISIMILITUDE

Catherine Elgin suggests, “to accept that p is take it that p ’s divergence from truth, if any, does not matter” (2004: 120) If the principle being a counterexample to itself is *irrelevant*, then the principle can continue to be accepted.

Acceptance does not imply belief. Belief is greedy, demanding full truth. Just try to believe a universal generalization (such as “All prime numbers are odd”) that has a single exception. You cannot do it. So perhaps the principle of verisimilitude is self-defeating with respect to belief but not with respect to acceptance.

What counts as a negligible departure from the truth may vary with purpose. So self-defeat may be correspondingly contextual.

25.19 EXEMPLIFICATION

Deviations from truth appear to count in *favor* of Xenophanes’ principle that we can only get close to the truth. We are not tolerating the error but somehow welcoming it. Nelson Goodman’s concept of exemplification offers a way of understanding this perverse welcome. He introduced “exemplification” to cover symbolic uses of a samples, as when a carpet swatch is used to refer to the texture and color of the carpet.

Exemplification is possession plus reference. To have without symbolizing is merely to possess, while to symbolize without having is to refer in some other way than by exemplifying. (Goodman 1976: 53)

Venus possesses hotness but does not refer to it. So Venus does not exemplify hotness. “Hotness” refers to the property of being hot but does not possess it. So “hotness” does not exemplify hotness.

Goodman notes that possession is an intrinsic relation but reference depends on a symbol system. The system establishes the referential relationship that “is a matter of singling out certain properties for attention, of selecting associations with certain other objects” (1976: 88).

Goodman further notes, “To exemplify or express is to display rather than to depict or describe” (1976: 93). The symbol highlights the targeted feature and thereby increases epistemic access to it.

Although denotation and exemplification are normally separate, they converge with self-referential terms:

A symbol that denotes itself also exemplifies itself, is both denoted and exemplified by itself. “Word” is thus related to itself, and so are “short” and “polysyllabic”, but not “long” and “monosyllabic”. “Long” is a sample of “short”, “monosyllabic” denotes short words, and “short” both exemplifies and denotes short words. (Goodman 1976: 59)

The convergence of self-reference and self-exemplification infects exemplification with Kurt Grelling’s variation of the liar paradox. If “heterological” exemplifies itself then it does not. And if it does not exemplify itself, then it does.

But Xenophanes might take this to be an appropriately blunted form of vindication.

REFERENCES

- Bradwardine, T. (2010). *Insolubilia*, ed. and trans. S. Read. Dallas Medieval Texts and Translations 10. Leuven: Peeters.
- Castagnoli, L. (2010). *Ancient Self-Refutation: The Logic and History of the Self-Refutation Argument from Democritus to Augustine*. Cambridge: Cambridge University Press.
- Cook, R. T. (2014). *The Yablo Paradox*. New York: Oxford University Press.
- Elgin, C. Z. (2004). “True Enough,” *Philosophical Issues* 14: 113–31.
- Ericsson, K. A. and J. Smith (eds.) (1991). *Toward a General Theory of Expertise: Prospects and Limits*. New York: Cambridge University Press.
- Garciadiego, A. (1992). *Bertrand Russell and the Origins of the Set-theoretic “Paradoxes”*. Boston: Birdkhauser.
- Geach, P. (1972). “Identity,” in Geach, *Logic Matters*. Oxford: Basil Blackwell, 211–43.
- Glanzberg, M. (2015) “Complexity and Hierarchy in Truth Predicates,” in T. Achourioti, H. Galinon, K. Fujimoto, and J. Martinez- Fernandez (eds.), *Unifying the Philosophy of Truth*. New York: Springer.
- Goldstein, L. (1992). “‘This Statement is Not True’ is Not True,” *Analysis* 52: 1–5.
- Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*, 2nd edn. Indianapolis: Hackett.
- Graham, A. C. (1989). *Disputers of the Tao*. La Salle, IL: Open Court.
- Haack, S. (1996). *Deviant Logic, Fuzzy Logic*. Chicago: University of Chicago Press.
- Herzberger, H. G. (1967). “The Truth-Conditional Consistency of Natural Languages,” *Journal of Philosophy* 64: 29–35.
- Herzberger, H. G. (1970). “Paradoxes of Grounding in Semantics,” *Journal of Philosophy* 67: 145–67.
- Honsberger, R. (1991). *More Mathematical Morsels*. Ann Arbor: The Mathematical Association of America.
- Kripke, S. A. (1984). “Outline of a Theory of Truth,” in R. L. Martin (ed.), *Recent Essays on Truth and The Liar Paradox*. New York: Oxford University Press, 53–81.
- Lu, P. et al. (2004). “Failure to Inhibit the Reading Response in Stroop Test: A Pathognomonic Indicator of Suspect Effort,” *Journal of Clinical and Experimental Neuropsychology* 26: 180–9.
- Lucas, J. R. (1961). “Mind, Machines and Godel,” *Philosophy* 36: 112–17.

- Maudlin, T. (2004) *Truth and Paradox*. Oxford: Clarendon Press
- Mills, E. (1998). "A Simple Solution to the Liar," *Philosophical Studies* 89: 197–212.
- Parfit D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Perrett R. (1984). "Self-refutation in Indian Philosophy," *Journal of Indian Philosophy* 12: 237–64.
- Prior, A. N. (1976). "Some Problems of Self-reference in John Buridan" in A. N. Prior, *Papers in Logic and Ethics*, ed. P. T. Geach and A. J. P. Kenny. London: Duckworth: 130–46.
- Popper, K. (1963). *Conjectures and Refutations*. London: Hutchinson.
- Post, J. F. (1970). "The Possible Liar," *Noûs* 4: 405–9.
- Quine, W. V. (1987). *Quiddities*. Cambridge, MA: Harvard University Press.
- Read, S. (1979). "Self-reference and Validity," *Synthese* 42: 265–74.
- Read, S. (2006). "Symmetry and Paradox," *History and Philosophy of Logic* 27: 307–18.
- Russell, B. (1905). "On Denoting," *Mind* 14: 479–93. Reprinted in Robert C. Marsh (ed.) (1956), *Logic and Knowledge: Essays 1901–1950*. New York: Macmillan, 39–56.
- Sorensen, R. (2001). *Vagueness and Contradiction*. Oxford: Clarendon Press.
- Sorensen, R. (2003). *A Brief History of the Paradox*. New York: Oxford University Press.
- Spade, P. V. and S. Read (2013). "Insolubles," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), <http://plato.stanford.edu/archives/fall2013/entries/insolubles/>
- Tarski, A. (1944). "The Semantic Conception of Truth: And the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341–76.
- Whitehead, A. N. (1953). *Dialogues of Alfred North Whitehead as Recorded by Lucien Price*. Boston: Little, Brown.
- Yablo, S. (1993) "Paradox without Self-reference," *Analysis* 53: 251–2.
- Yablo, S. (1985). "Truth and Reflection," *Journal of Philosophical Logic* 14: 297–349.

CHAPTER 26

TARSKI ON THE CONCEPT OF TRUTH

GREG RAY

ALFRED Tarski's work on truth has played such a central role in the discourse on truth that most coming to it for the first time have probably already heard a great deal about what is said there. Unfortunately, since the work is largely technical and Tarski was only tangentially philosophical, a certain incautious assimilation dominates many philosophical discussions of Tarski's ideas, and so, examining Tarski on the concept of truth is in many ways an act of unlearning.

In the following, I will focus on those key ideas in Tarski's work that have made a lasting impact on the philosophical discourse. These are the notions of T-sentence, Convention T, Tarskian truth definition, and Tarski's general limiting theses on the expressibility and definability of truth. Though these ideas are in name familiar, we will seek in this chapter to uncover and remove certain widespread misunderstandings of each. Tarski's name also features prominently in discussions of the liar paradox, so we will take time out to explain Tarski's connection to this ancient puzzle.

26.1 TARSKI'S PROJECT

In his first, central essay on truth, Tarski (1935b) tells us he has a single goal—the construction of “a materially adequate and formally correct definition of the term ‘true sentence.’”¹

¹ Also, Tarski (1944: 341). Today, “materially adequate” has been largely co-opted as a technical term for something which meets Tarski's adequacy condition, Convention T. But Tarski's own use of this phrase—which occurs in these passages without introduction—is one he takes to have an accepted or immediately grasped meaning. I hazard to suggest that both Tarski's original Polish phrase “*merytorycznie trafna*” (1933) and his German translation of it as “*sachlich zutreffende*” (1935a)—which has commonly been translated as “materially correct”—might as happily be translated as “factually

Semantic paradox—especially the liar paradox—is often thought of as a prime motivation for Tarski's search for such a definition. However, Tarski tells us at the outset that a “more or less satisfactory” solution to these paradoxes had already been found.² The avoidance of paradox is an essential first step in his investigation, to be sure, but Tarski took his own contributions to be i) a novel solution to the problem of generality which beset earlier attempts to define truth; and, by virtue of offering the first logically precise truth definition, also ii) proof that there are definite limits to the expressibility and definability of truth.

Tarski saw his work on truth as an advance on earlier efforts to characterize truth such as Aristotle's:

To say of what is [so] that it is [so], or of what is not [so] that it is not [so], is true.

as well as formulations such as

a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so. (Tarski 1935b: 155)

As such, Tarski is on the track of what we may call the “correspondence intuition.” However, Tarski is not a classical correspondence theorist. His strategy of truth definition features no correspondence relation and lacks entirely its associated reification of facts or states of affairs. In fact, it is one of the signal accomplishments of Tarski's proposal that it shows how one can give a general characterization of a truth predicate that is genuinely semantic and in accord with the correspondence intuition, but does not resort to reification.³

Tarski shared with contemporaries Russell and Ramsey the sense that the problem in defining truth is a *problem of generalization*. All three opined that what truth comes to is easy to say for particular cases, but does not have a straightforward generalization.⁴ Generalizing in the most obvious way yields ungrammatical constructions (Ramsey 1929 [1991]: 9–10) or ontological posits (Russell 1912) not evident in the original. It was this problem of generality that vexed Russell and Ramsey, and for which Tarski, using the then-new technique of inductive definition, had a solution.⁵

accurate” or “factually correct.” Thus, the aim would be a definition which was both factually and formally correct.

² Tarski (1935b: 152). Tarski may refer here to Russell's ramified theory of types or perhaps more likely to the view of his own teacher, Stanisław Leśniewski, whose theory of semantic categories Tarski clearly favors (1935b: 215). See also Luschei (1962: 35).

³ There is a history of debate on whether to count Tarski a correspondence theorist. Cf. Kirkham (1992: 170) for references. Kirkham's own discussion is unfortunately marred by a hasty attribution to Tarski of a sort of underlying view which reifies states of affairs (1992: 167). At some points, Kirkham seems to take Tarski's T-sentences as making veiled reference to states of affairs—apparently construing the right side of such biconditionals as referring terms (1992: 172).

⁴ Cf. Tarski (1935b: 158–60).

⁵ Soames (1999: 69).

Speaking of doxic truth, Ramsey put the problem in this way:

Suppose a man believes that the earth is round; then his belief is true because the earth is round; or generalizing this, if he believes that A is B his belief will be true if A is B and false otherwise. It is, I think, clear that in this last sentence we have the meaning of truth explained, and that the only difficulty is to formulate this explanation strictly as a definition. If we try to do this, the obstacle we encounter is that we cannot describe all beliefs as beliefs that A is B since the propositional reference of a belief may have any number of different more complicated forms. A man may be believing that all A are not B, or that if all A are B, either all C are D or some E are F, or something still more complicated. We cannot, in fact, assign any limit to the number of forms which may occur, and must therefore be comprehended in a definition of truth; so that if we try to make a definition to cover them all it will have to go on forever, since we must say that a belief is true, if supposing it to be a belief that A is B, A is B, or if supposing it to be a belief that A is not B, A is not B, or if supposing it to be a belief that either A is B or C is D, either A is B or C is D, and so on ad infinitum. (Ramsey 1929 [1991]: 9)

This is the problem that Tarski will address—for the case of sentential truth in particular.⁶

What sort of definition did Tarski aim to offer and what does it mean for it to be formally correct or materially adequate? Being *formally correct* just means meeting what are now quite standard formal rules of definition which ensure that a definition is eliminative and non-creative.⁷ The former guarantees that the defined predicate can always be eliminated by putting in its place the defining phrase. This constraint effectively rules out circular definitions. Non-creativity ensures that the definition does not unwontedly make new theorems of any theses but ones which use the defined term.⁸ However, “true” is not a logical term and no *merely formal* constraint can fully ensure that we have the desired definition. So, by *material adequacy* (*factual correctness*), Tarski just means a definition that is (in some sense) appropriate given what this non-logical word in fact means. Famously, Tarski proposed his Convention T as a precise measure of the factual correctness of the sort of definition he would propose.⁹

⁶ Tarski assumed that a solution for sentential truth would by extension also be a solution for things like doxic truth—on the strength of the idea that claims about true beliefs can (in some sense) be rephrased as claims about true sentences.

⁷ Stanisław Leśniewski, who was Tarski’s dissertation advisor, held that many logical systems came to grief for not having been careful in setting out rules of definition, and apparently made a small industry of showing how this was so in various systems. Leśniewski had already formulated quite precisely formal rules of definition for various systems that are equivalent to the standard rules of definition adopted in standard classical logic texts, including Tarski’s (1936a). Cf. Luschei (1962: 36–7).

⁸ For a standard exposition, see Mates (1972: 199–202).

⁹ The notion of “material adequacy” is sometimes simply identified with Tarski’s Convention T, but as noted earlier Tarski evidently used material adequacy in a more commonsensical way—the notion appears at the very outset of Tarski (1935b) well before Convention T is formulated. Rather, Convention T is meant as a precise way of ensuring material adequacy—and we are to see that it does so by our antecedent grasp of the concept of truth. It is worth noting that material adequacy is also conceptually distinct from *extensional correctness*—for the reason that there are languages for which a truth predicate can have no well-defined extension. A suitable condition on material adequacy should imply extensional correctness, but ought not presuppose extensional existence.

What sort of definition should we think of Tarski as offering? A number of philosophical issues have been thought to hang in the balance and effect our assessment of Tarski's achievement.¹⁰ The question is typically framed as a choice between two types of definition: merely extensionally correct definition or meaning-giving definition.¹¹ This is not a particularly helpful approach, however. While it is clear that Tarski's notion of definition is just the formal logical one and there is nothing in that which requires the capturing of meaning, and it is equally clear that a materially adequate definition of a non-logical term does not require more than (roughly speaking) extensional correctness, still neither of these points implies that Tarski does not have an interest or a reason to provide more.

An important feature of any strategy Tarski might pursue would be that we be able to *secure conviction* that a resulting definition was indeed materially adequate. It is by our understanding of the meaning of the term "true" that we are able to see that a definition meeting Convention T would be "factually accurate," and it is this that recommends Convention T to us.¹² Understanding this already goes some way toward reconciling various passages in Tarski which might otherwise appear in tension with each other as between a commitment to merely extensional definition and talk of "capturing the meaning" of the truth predicate.

In our discussion of Convention T, we will see that Tarski is committed to meaning over and above extension—and in a way quite unfriendly to disquotationalism.

26.2 THE T-STRATEGY

The notion of T-sentence and the principle, Convention T, which it serves form the basis of Tarski's T-strategy for defining sentential truth and solving the problem of generality. Tarski's teacher, Leśniewski, had drawn Tarski's attention to T-sentences as playing a role in formulations of the liar paradox. It was Tarski's idea that the general class of such sentences, if they could be comprehended in a finite definition, could form the proper foundation of a truth definition (Tarski 1944: 343 fn. 7).

A *T-sentence* is a translational T-form sentence, i.e., a sentence that is i) of the form "s is true iff p" where "s" is replaced by a name for an object language sentence, and

¹⁰ The point has for example been pressed that Tarskian definitions are not meaning-giving (Putnam 1983: 37–40; 1985: 63–4; Etchemendy 1988: 56–7; Soames 1995: 253–4). On this basis, John Etchemendy e.g. argued that Tarski cannot have contributed to theoretical semantics as has been thought (Tarski 1936c). For some critical discussion of such arguments, see Davidson (1990: 288–95); García-Carpintero (1996); Heck (1997).

¹¹ Etchemendy (1988); Heck (1997); Kirkham (1992: 20, 22, 37); Haack (1978).

¹² In Tarski (1935b: 155, 157) this recognition is rooted in our seeing that T-sentences are meaning-instance giving. This same suggestion is not repeated in Tarski (1944: 343). But one should be cautious in taking talk of meaning in Tarski in full philosophical seriousness. Consider the strikingly loose usage in Tarski (1969: 64) where it is said that "by saying that S is true we mean simply that snow is white."

“p” is replaced by a sentence of the theory language, and ii) the sentence replacing “p” is a translation of (means the same as) the sentence referred to by the name which replaces “s”.

It is important to realize that not all well-formed T-sentences need be true—even if we set aside possible cases arising from the use of vague terms, for example. That not all T-sentences need be true is an immediate consequence of an argument that Tarski makes strategic use of—in fact, the very argument that brought T-sentences to Tarski’s attention. If S is a liar sentence—“S is not true”—which says (in language L) of itself that it is not true (in L), then a T-sentence for S in L is

“S is not true” is true iff S is not true

Tarski shows that this T-sentence can be used to derive a contradiction with the help of the uncontroversial factual claim that

S is the sentence “S is not true.”

Prima facie, the argument is an obvious *reductio* of the above T-sentence. Nonetheless, a surprising number of philosophers seem to presuppose that all T-sentences are true, and sometimes take this as a central pillar of the Tarskian framework.¹³ Part of the fault for the latter is Tarski’s. First, because at what seems the opportune moment, he does not explicitly draw the obvious *reductio* conclusion from the liar argument as above. As it happens, his direct aim in displaying that argument—namely, to show that a satisfactory truth definition in L for L is not possible—does not require his making the current point, but neither does it require that he be committed to the truth of the T-sentence premise of the argument.¹⁴

Second, many readers when first coming to Tarski’s introduction of T-sentences naturally read him as suggesting that these sentences be accorded a special role in the proceedings because of their evident truth. And indeed, the simple example Tarski famously gives (for “snow is white”) is quite evidently intended as unproblematically true. However, analysis of the relevant texts confirms (the hardly surprising fact) that Tarski did not think of all T-sentences as true. I will not attempt to rehearse the case for this here, but a leading observation would be to notice that T-sentences are initially characterized only as sentences of a special kind which present or recommend

¹³ As a cautionary example, we can note that Paul Horwich’s original (1990) made a significant investment in making all disquotational T-propositions true. In a later revision (Horwich 1998), this is amended in the face of the fact that they would together form an inconsistent set. T-sentences have variously been thought to be either conceptual, necessary, or analytic truths. None of these is possible, of course, if some T-sentences are not even so much as true. Cf. McGee (1992).

¹⁴ A definition that met Convention T would have to validate the main premise of the above *reductio* argument. This fact essentially shows that such a definition cannot be formally correct, because it is creative. This was Tarski’s point in a nutshell. We will say a bit more about this issue later.

themselves to us (Tarski 1935b: 155). And of example T-sentences such as his famous “snow is white” case Tarski only says they “seem to be clear and completely in accord with the meaning of the word ‘true’ . . . In regard to the clarity of their content and the correctness of their form they arouse, in general, no doubt” (1935b: 157). In light of our question, these passages are striking for what they omit. Likewise, one finds in crucial formulations where one might expect an attribution of truth, Tarski speaks instead of languages in which T-sentences are *considered* or *regarded as* true.¹⁵

Since T-sentences and Convention T which is based on them have a significant standing in certain areas of philosophy, it is important to appreciate this point.¹⁶ The *reductio* argument above gives us the strongest sort of reason to expect that in expressive natural languages, not all T-sentences need be true. Nonetheless, T-sentences do, as Tarski says, *recommend themselves to us* and have an important role to play in understanding the concept of truth. Even though T-sentences cannot be a class of analytically true sentences—as one might justifiably expect from Tarski’s plan to “build them into” a truth definition—there is nonetheless a tight conceptual connection between T-sentences and our concept of truth.^{17,18}

In Convention T, Tarski formulated a precise success condition which succinctly expresses his T-strategy of truth definition.

Convention T is the requirement that the proposed definition make theorems of a complete set of T-sentences for the language in question.¹⁹ A complete set, C, of T-sentences for an object language is such that each sentence of that language is referred to by some T-sentence in C.²⁰

¹⁵ For a fuller examination, see Ray (2003).

¹⁶ Considerable mischief can ensue. Scott Soames (1999: 52–3) attributes to Tarski the view that all T-sentences are true, and it is almost certain that this early mistake helps construct the dilemma that later pushes Soames to the logically confounded view that some sentences (e.g. liar sentences) must be *neither true nor not true*. Cf. Ray (2003: section 2).

¹⁷ Scott Soames makes the point in (1995)—which is an interesting discussion of the status of T-sentences. For an explicit characterization of the special status of T-sentences, see Ray (2002). To put it roughly and somewhat mysteriously, T-sentences can be seen to have going for them just what conceptual or analytic truths do, but for the determination of their truth value.

¹⁸ Tarski’s idea of taking T-sentences as “partial definitions” of truth and constructing a definition that will be the “logical sum” of these is complicated for a little-observed reason. There is at least one salient case in which such a “partial definition” must be improper because circular, namely in any “true”-involving case. For an example, consider: “*Knemon is happy*’ is true” is true iff “*Knemon is happy*” is true. While the claim may be unproblematically true, it is not unproblematic to treat it as an instance-definition of “true.” Tarski recognized the case (1935b: 158). But he appears not to see a further problem in this, presumably since languages in which such cases arise will be—for a different reason—ruled outside the realm of his proposed definitions.

¹⁹ Interestingly, Tarski formulates Convention T with a narrower class of T-sentences, namely those which employ a so-called structural-descriptive name for the target sentence, as opposed to any other name. It is left as an exercise for the student to determine what ground there could be for this as a requirement of factual correctness. Tarski offers none.

²⁰ Note that meeting Convention T does not require making theorems of *all* the T-sentences, as many formulations imply. We only want to insist on a complete set of T-sentences. These demands

The first thing to observe is that, contrary to popular belief, Convention T is formulated by Tarski as a success condition for a particular strategy of truth definition—his. Tarski does not allege that *any* approach to truth definition whatsoever must cleave to this condition.²¹ This is just as well, since we will see in section 26.6 a clear example of an extensionally correct truth definition that nonetheless does not satisfy Convention T.

Second, what we earlier observed about T-sentences raises a challenging but little-noticed problem. If not all T-sentences need be true, why would we insist that a truth definition make theorems of a complete set of these T-sentences—for this is precisely what Convention T insists that we do. Doesn't this just make Convention T look like a bad measure of adequacy?

Consider how Tarski argued that one could not define true-in-English in English. The reasoning went something like this: such a definition would make theorems of a complete set of T-sentences, but since English is expressive enough to formulate in it liar sentences, among these theorems would be the T-sentence for such a liar sentence. By the foregoing argument, a contradiction could then be derived, thereby demonstrating, contrary to hypothesis, that the definition on offer was creative and hence not formally correct.

Notice how this line of reasoning depends on the idea that any satisfactory truth definition would have to sanction (make theorems of) a complete set of T-sentences—even the ones which we know cannot be true. But rather than deciding this constraint is a bad or at least over-zealous one, Tarski holds to the constraint and concludes instead that there could not be a satisfactory truth definition (for English in English).²²

We can make sense of this judgment only by understanding Convention T as rooted squarely for Tarski in the *meaning* and not the extension of the truth predicate. And this is surely how it should be because, as hinted earlier, semantic paradoxes make it problematic to think of the truth predicate in many languages as having a well-defined extension in the first place. This is the sense in which Tarski makes good on his claim to “capture the meaning.” It is not that a Tarskian definition itself is meaning-giving. Rather, it is because the shape of a Tarskian definition is constrained by a requirement that is fixed by the meaning of the target term—even at the expense of projecting on the predicate an impossible extension. Indeed, as we will see in section 26.4, it is part and parcel of Tarski's underlying view that, for a peculiar logical reason, when it comes to the truth predicate, meaning and extension must come apart.²³

differ because there is in general no reason to expect that only one T-sentence for a given sentence will be expressible. This comports well with Tarski's introduction of Convention T (1935b: 187–8) where a particular complete set of T-sentences is referenced—by there specifying a particular, unique T-sentence for each sentence (1935b: 172).

²¹ Though Tarski does argue that no right-thinking definition could get results genuinely *opposed* to it. This, on the grounds that that would require an extensional difference that the opposing definition would be *perforce* on the losing end of (Tarski 1944: section 18).

²² Gupta and Belnap (1993: 12); Ray (2002).

²³ For this reason, it would seem a mistake to classify Tarski as a deflationist, as many have done.

26.3 THE LIAR

Tarski's salient use of a liar argument has given him a position in discussions of the liar paradox, and it is commonly thought that Tarski offered some sort of solution to the paradox. To be clear, let us distinguish in the usual way between *diagnoses* of the liar and *directives* for its avoidance²⁴—both of which are sometimes called “solutions.” Put in these terms, Tarski definitely offered a well-known directive for paradox avoidance. He tells us that we must clearly distinguish the *metalanguage* in which a truth definition is offered from the *object language* whose sentences are the range of application of that predicate, and he identified conditions on these languages which were meant to be sufficient to avoid paradox.

When philosophers speak of Tarski's solution to the liar, they typically have in mind his directive—and some at least have mistaken it for a diagnosis. Now, such a directive is not in itself a diagnosis of the paradox. Its aim is fundamentally practical and shows us how to get on with the business of truth predication while avoiding self-contradiction. For the pursuit of his truth definition project, this is all Tarski requires and all he offers.

So, Tarski did not explicitly offer any diagnosis of the liar at all and while it is possible to reconstruct something of a diagnosis from his texts (see section 26.4) and the resulting view is acknowledged in various places in the literature, this is almost never what philosophers speak of when they talk about Tarski's solution to the liar.

Rather, when philosophers speak of Tarski's solution to the liar paradox, what they typically have in mind are salient passages in which Tarski first identifies certain conditions on languages that, when met, preclude the possibility of pursuing his T-strategy due to liar-like pathology, and then proposes to restrict his subsequent inquiry to languages that do not meet those trouble-making conditions. Since the barrier to definition here is related to the liar paradox, Tarski's proposal is clearly what we called a *directive* for paradox avoidance. Still, philosophers have sometimes mistaken this for something more substantive. This misunderstanding is partly Tarski's fault.

In (1944: section 7), after presenting the liar argument, Tarski tells us that an antinomy like this presents us with an intolerable situation in the face of which we cannot rest content but must discover its cause—reject a premise. *But he does not offer to do so.* That section of the paper ends and he picks up in the next section in a way that has led many to suppose that they were hearing Tarski's own pronouncement on the “intolerable situation”—that is, what was wrong with the liar argument and which premise of it was to be rejected.

If we now analyze the assumptions which lead to the antinomy of the liar, we notice the following:

- (i) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well

²⁴ Relevant distinctions are made in Chihara (1979) and Gupta and Belnap (1993: 10–11).

as semantic terms such as the term “true” referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called “semantically closed.”

- (ii) We have assumed that in this language the ordinary laws of logic hold.
- (iii) We have assumed that we can formulate and assert in our language an empirical premise such as the statement (2) which has occurred in our argument.

It turns out that the assumption (iii) is not essential, for it is possible to reconstruct the antinomy of the liar without its help. But the assumptions (i) and (ii) prove essential. Since every language which satisfies both of these assumptions is inconsistent, we must reject at least one of them (1944: 348–9).

It seemed as though Tarski was setting up to tell us what premise of the liar argument was to be rejected, but this is not what he does. It can't be. The assumptions he lists here are *not* the premises of the liar argument at all, and the “rejection” to be offered here is not a rejection of the expected sort.

What Tarski does is *choose* to only consider languages that are not *semantically closed*, i.e. that do not meet condition (i). But no one could sensibly think Tarski is *rejecting* (i) in the sense of saying there are no semantically closed languages. He clearly thinks that natural languages are a case in point. What is going on? Tarski is just thinking of these conditions as ones which jointly prohibit a satisfactory truth definition being given. In short, he is not focused here on saying anything about the liar per se. He has moved on to the task of determining a restricted context in which the problem which he has set himself may yet be solvable. For these purposes he is only seeking to avoid the liar, not to assess it, diagnose it, or explain it. No doubt it is in part this unfortunate segue in Tarski (1944) that has misled many philosophers into thinking that Tarski tried to solve the liar paradox in the diagnostic sense (Ray 2003).

I say it is partly Tarski's fault that his directive was sometimes taken for a diagnosis, but I do not think it is entirely Tarski's fault. Philosophical readers are perhaps too quick to assume that any presentation of a liar argument must be propaedeutic for an impending solution, and so have paid insufficient notice to the precise use to which Tarski puts his liar argument. Understanding this is the key to not *misunderstanding* what Tarski does and does not have to say about the liar.

Recall Tarski aimed to solve the generality problem but that his T-strategy for this met a particular challenge traceable to the semantic paradoxes.²⁵ In his *informal indefinability argument* he argued that the T-strategy of truth definition cannot be carried out in natural language (or any language very like it). It is to establish this result that Tarski made reference to a liar argument, and this is the only role that the liar plays in Tarski's project.²⁶

²⁵ In Tarski's preferred manner of presentation (1935b: 165 fn. 1), it is actually the Grelling that bites, not the liar. For an analysis of this difference, see Ray (2006).

²⁶ The liar argument is formulated at 158 but the lesson postponed to the argument of 164–5 of Tarski (1935b).

We briefly alluded to the reasoning of Tarski's informal indefinability argument earlier. Suppose you have a definition of truth for English in English, and suppose this definition meets Convention T. The definition must then make theorems of a complete set of T-sentences for English. If all the T-sentences for English are theorems, then LA, the liar-based argument presented in section 26.2, would have one premise which was (now) a theorem and the other a simple identity premise.²⁷ In the particular version of the liar argument that Tarski gave, the second premise is a clearly empirical one and Tarski observes that we can give a version of the argument which would eliminate dependence on any such premise.²⁸ It would follow that the conclusion of LA must itself be a theorem and the definition will have made it so. But argument LA is an argument to contradiction and a definition that enables us to derive contradictions is a creative definition and hence one that is not formally correct.²⁹

In its proper context, the liar argument LA is merely referenced in Tarski's own argument, where it is used as leverage to obtain an initial indefinability result—a result intended to show that the T-strategy has no hope of success in/for natural languages (or languages relevantly like them). So far, then, Tarski's discussion and use of the liar yield no proffered analysis.

26.4 TARSKI'S THEORY OF TRUTH

Still and yet, though it is not Tarski's focus and we get no pronouncement of the fact, his position does *suggest* a diagnosis of the liar paradox. His commitment to classical logic together with his view that liar sentences are not "improper,"³⁰ together imply for us that the liar argument is just a *reductio* of its T-sentence premise. Hence, the T-sentence for a liar sentence is just not true. This, together with Tarski's evident commitment to the T-strategy

²⁷ To repeat, the two premises are the T-sentence of a given liar sentence, i) "L is not true" is true iff L is not true, and a premise which identifies the liar sentence that L is, ii) L is the sentence "L is not true."

²⁸ To really make this work, Tarski means to use a Grelling argument, not a liar argument, in fact. For discussion, see Ray (2006). To cut to the chase, we have here used a version of LA that makes the premise already non-empirical in Tarski's sense, but notice that the result is at best a semantic truth, not a logical one as Tarski seems strictly speaking to need in order to implicate the proposed definition as one which is not formally correct—as opposed to in some lesser sense an unhappy one.

²⁹ I have set aside an aspect of Tarski's approach which has occasioned a good deal of confusion and drawn attention away from Tarski's main point. In the event, Tarski's way of establishing his result is by first drawing the conclusion that the language in question must be inconsistent, and then from this concluding that no formally correct truth definition in that language is possible. The middle step is obviously unnecessary and many have been (for some reason) confounded at the notion of an inconsistent language. To say that an exactly specified language is an inconsistent language is just to say that the set of theorems of the language is deductively inconsistent—and Tarski has in mind some cognate notion for natural languages. For a full discussion, see Ray (2003: section 1).

³⁰ Of the T-sentence for a liar sentence, Tarski holds that "no rational ground can be given why such [sentences] should be forbidden in principle" (1935b: 158).

even with respect to natural languages (in which such T-sentences arise) fairly implies that the concept of truth is *incoherent*—it is assigned the task of expressing a concept of truth (and this determines its meaning), but the dictates of that concept do not entirely cohere with each other (and this is manifested in its extensional pathology). To hold this position is to answer both of the central questions that a diagnosis of the liar requires, namely, to tell us what premise to give up in the liar argument and to explain what the underlying problem is (Chihara 1979: 590–2).

The incoherence view of truth has been a rather unpopular one, but Tarski's commitments clearly commit him to it.³¹ Getting to this result requires some extrapolation from what Tarski says, not because we have needed to add something essential of our own nor because the implication is uncertain, but only because Tarski's project required no pronouncements on this question and so as a matter of fact he did not make any.

It is worth expanding on this idea a bit, since if Tarski has a theory of the concept of truth per se, this is central to it. By dint of our linguistic intentions and practices the meaning of a predicate "true" is fixed—it comes to express a certain concept, the concept of truth. In this way we make it the case that the word "true" is *supposed* to work a certain way—it is supposed to *apply* to certain things and it is supposed to *fail-to-apply* to certain other things.³²

It is clear that Tarski holds the common-sense view that the English phrase "true sentence of English" does express our concept of sentential truth for English,³³ and Convention T is designed to constrain a definition to respect the assignment of meaning this implies. In so doing, however, it forces any proposed definition to imply an untruth³⁴ which in turn can be used to infer contradictions. In short, if this truth predicate did all we have unwontedly *meant* for it to do, it would have to both apply *and* fail to

³¹ It has been suggested to me that philosophers who find the incoherence view unpalatable will find in this reason enough to doubt the attribution and cause for finding a reading of Tarski that avoids the implication. I wish any honest scholar good luck with that, but it seems to me that one's own distaste for the view will not justify going to any lengths. It is simply not at all implausible that Tarski would hold the view, and, moreover, if we are to eschew "finesse" and cleave to the textual evidence, it appears a settled matter that he held things that commit him to it. On the first point, we need only remark that the view seems natural kin to the conclusion standardly drawn in regard to our naïve concept of collection (set) in the face of set-theoretic paradox, namely that our ordinary notion of collection is in a certain way incoherent. One hardly defames Tarski by suggesting he held the corresponding view with regard to our concept of truth in the face of truth-theoretic paradox. On the second point, I can do no better here than ask the skeptical reader to consider again carefully the solid grounds of attribution given in the first paragraph of this section.

³² In the terminology of Ray (2002), we assign the predicate the job of *subserving* the concept of truth. Using the useful fiction that concepts come outfitted with application conditions, this is just to say that if the application conditions imply that a thing falls under the concept, the predicate is intended to apply to that thing, and if the application conditions imply that a thing fails to fall under the concept, then the predicate is intended to fail-to-apply to that thing. We do not presuppose here that applying and failing-to-apply are jointly exhaustive. For more on this important distinction, see also Ray and Ludwig (2002).

³³ This is an unquestioned assumption of most discussions of the liar paradox, but one which ought not go unexamined. Cf. Badici and Ludwig (2007).

³⁴ One of the T-sentences that is not true.

apply to certain things—such as liar sentences. But this is just to say that the predicate has been given a job (subserving our concept of sentential truth) that it cannot possibly perform—the extension that would be required is impossible.

It is for this reason that we should count Tarski as an *incoherency theorist* about truth. It is hard to see how cleaving to Convention T can be justified in light of its effect other than as an admission that the dictates of the concept of truth do not cohere, and this is manifested in the apparent pathology of the truth predicate that the liar paradox famously brings to light.³⁵

26.5 TARSKIAN TRUTH DEFINITION

We turn now to Tarski's actual method of truth definition—an inductive definition which will meet Convention T and thereby in a precise way solve the problem of generality. In addition to its obvious impact on the philosophical discourse on truth, it would be genuinely hard to overestimate the impact that Tarski's technique has variously had in logic, and in philosophy of language and linguistics, as well as in many other areas of philosophy.

In some sense, the general shape of a Tarski-style truth definition is known by everyone who has studied the formal semantics of classical logic—wherein one defines for a semi-interpreted language, *L*, the notion of *truth-in-an-L-structure*, or *truth-under-an-L-interpretation*.³⁶ The modern presentation of these notions is a development from Tarski's (1936b) work on logical consequence which is itself founded on his work on truth definition. However, one schooled only in the model-theoretic definitions of truth and satisfaction, and who has not already encountered in some cognate area the idea of an absolute truth definition for a fully interpreted language—for example in philosophy of language—will not yet be able to properly appreciate the philosophical discussion surrounding this sort of truth definition.

Fortunately, examples of Tarski-style truth definitions abound in the literature and exist in varying levels of detail. This makes our current task easier—since we can leave many details to more extended discussions. Here we will simply use the original example given by Tarski—simplifying and eschewing unnecessary notation wherever possible.

Tarski gave a definition of *true sentence in the language of the calculus of classes*. Everyone who has studied formal logic knows how the theory of syntax for a simple language is built up. The basic vocabulary of the language of the calculus of classes includes

³⁵ Chihara (1979; 1984). The current consideration suggests that many philosophers who accept Convention T but would not accept the incoherence thesis are in a rather delicate position.

³⁶ By *semi-interpreted*, I just mean a language where the only terms with intended interpretations are logical terms. Semi-interpreted languages are standard in basic logic courses, because for the purposes of formal logic, the non-logical terms of the language in question need have no antecedently intended interpretations.

a two-place relation symbol for inclusion (i.e. subset), logical symbols for negation, disjunction, and universal quantification, plus an inexhaustible collection of variables for binding. Definitions of the requisite syntactic notions—bound and free variables, atomic, molecular, and quantified formula and sentence—are completely canonical.

To define Tarski's notion of *satisfaction* we need one notion that is sensitive to the *intended interpretation* of our object language. On that interpretation, the language of the calculus of classes has in its domain only the subclasses of a fixed set of ur-individuals.³⁷ So, the notion we need is that of an infinite sequence of these classes—which we will call here a *c-sequence*. We can now give the inductive definition of satisfaction:

For all *c*-sequences *s* and formulas *p*, *s* *satisfies* *p* if and only if

- a) *p* is an inclusion formula in the *n*th and *m*th variables, and the class which is the *n*th item of *c*-sequence *s* is a subset of the class which is the *m*th item of *s*, or
- b) *p* is the negation of a formula *q* and *s* does not satisfy *q*, or
- c) *p* is the disjunction of formulas *q* and *r* and either *s* satisfies *q* or *s* satisfies *r*, or
- d) *p* is the universal quantification of *q* and binds the *n*th variable therein and every *c*-sequence *s*^{*}, which differs at most in the *n*th position from *s*, satisfies *q*.

Truth for this object language is then defined in terms of satisfaction.

For all formulas, *p*, *p* is *true* if and only if i) *p* is a sentence, and ii) for all *c*-sequences *s*, *s* satisfies *p*.

In Tarski's original understanding, inductive definitions were not themselves proper, but only stand-ins for explicit definitions. To obtain an explicit definition, we can replace our inductive definition of satisfaction with an explicit definition.³⁸

A *c*-sequence *y* *satisfies* a formula *z* if and only if $\langle y, z \rangle$ is an element of the least set, *X*, such that, for all *c*-sequences *s* and formulas *p*, $\langle s, p \rangle$ is in *X* if

- a) *p* is an inclusion formula in the *n*th and *m*th variables, and the class which is the *n*th item of *s* is a subset of the class which is the *m*th item of *s*, or
- b) *p* is the negation of a formula *q* and $\langle s, q \rangle$ is not in *X*, or
- c) *p* is the disjunction of formulas *q* and *r* and either $\langle s, q \rangle$ or $\langle s, r \rangle$ is in *X*, or
- d) *p* is the universal quantification of *q* and binds the *n*th variable therein and every *c*-sequence *s*^{*}, which differs at most in the *n*th position from *s*, is such that $\langle s^*, q \rangle$ is in *X*.

Thus, the use of the method of inductive definition is not essential to Tarskian truth definition.³⁹ However, it is to be noted that this explicit definition of satisfaction is of

³⁷ In Tarski's type-theoretic set-up, this is a language of order one. These sets will stand in inclusion relations to each other, but none will be a member of any of the others. They do not form a hierarchy of sets. No sets of sets at all. Moreover, the ur elements of the sets are not themselves referenced by any term of the language.

³⁸ Cf. Tarski (1935b: 193 fn. 1).

³⁹ Cf. Tarski (1935b: 177 fn. 1). The point is made by Wilfrid Hodges (2004: 102, 104).

higher-order than the inductive definition—since it must quantify over not only sequences and formulas, but also sets of pairs of these (or over relations between them).⁴⁰ This issue exercised Tarski a good deal as he turned to consider more expressive language systems. It is part of the reason (though not an ultimate reason)⁴¹ that he came to think that the metalanguage in which one defines a truth predicate must be essentially richer than its target language, and hence that there were definite limits to the definability of truth.

For a closer view of Tarskian truth definition, Tarski's own example definitions (22 and 23) of satisfaction and truth for the *calculus of classes* is to be found in Tarski (1935b: 193–4), but gaining the necessary understanding of the previous twenty-five pages with their antiquated notation can be a burden. Hence our brief sketch of his example above. A good detailed example can be found in Soames (1999: 71–81), and an informal discussion of Tarski's basic strategy of definition can be found in Kirkham (1992: 144–62).

26.6 LIMITS OF TRUTH

To Tarski's work on truth we can trace several general limiting theorems or theses that have had a significant impact on discussions of truth. One of these is his *informal indefinability thesis* which we have already discussed. We can distinguish three others: a) the *inexpressibility theorem* (commonly known as *Tarski's theorem*); b) the *formal indefinability theorem*; and c) the *essential richness thesis*.

Stated most crudely, these are the theses that i) a logical system⁴² cannot “have its own truth predicate”, i.e. cannot express its own truth set,⁴³ ii) a logical system cannot

⁴⁰ The explicit definition has a liability that the inductive definition does not, namely, that it relies on a non-syntactical notion that is not already present in the object language. Hence, it cannot be said for the explicit definition—like it can be for the inductive definition—that it is as safe as the object language, because truth claims are eliminable in favor of (cognates of) terms of the object language together with innocent terms of syntax. The explicit language could not do better than elimination in favor of cognates plus syntax plus set theoretic talk. So, the explicit definition has a bit of ontological burden the inductive definition does not. Davidson (1990) made the point, though his interest was the fact that an axiomatic theory of truth held the same advantage as the inductive definition. The result is that the explicit definition is only as safe as the underlying set theory that is employed. Thus, allowing inductive definitions, as themselves legitimate definitions as Tarski does in (1944), has a further tangible upshot: Tarskian truth definitions are just that little bit safer for not having to rely on the notion of set. This may seem negligible to those raised on modern axiomatic set theory, but taken in the light of the history of set theory may not be so small as it seems.

⁴¹ For reasons explained in Ray (2005: section 1.2).

⁴² By this I mean what Tarski called an “exactly specified language”—a syntactically specified, fully interpreted language together with a distinguished set of axioms, rules of inference, and rules of definition. I will rely elsewhere on the well-established expedient of referring to this as a *language system* or, more briefly, a *language*.

⁴³ An interpreted language *expresses a given set* just in case there is an open formula in one free variable which has the given set as its extension, i.e. is satisfied by all and only the elements of that given

define its own truth predicate, and iii) for a metalanguage system to define a truth predicate for a target language system it is necessary and sufficient that the metalanguage be essentially richer than the target language. None of these theses should be accepted as such, but these statements reflect in a rough and ready way the popular gist of Tarski's general theses. In the following, we will discuss each of these theses in turn. Our discussion will necessarily be brief, but we will aim to dispel several widespread misunderstandings.

- *The Inexpressibility Theorem.* A certain sort of formal result, sometimes called Tarski's theorem, is often mentioned in discussions of truth. Different philosophers say different things about what the significance of the result is, but all seem to agree that it is of some certain significance for our philosophical understanding of the concept of truth.

The formal result I have in mind here is characteristically proved in a way that uses some Gödel coding and something like the diagonalization lemma—in fact, the result is sometimes presented as an easy companion to Gödel's first incompleteness theorem.⁴⁴

According to its general acceptation, the theorem tells us that a language system of moderate theoretical strength cannot express its own truth set. We must distinguish this *inexpressibility* result from Tarski's *indefinability* thesis. The former result is about whether a language system can express its own truth set, whereas the latter turns out to be about whether a metalanguage could have a Tarskian truth definition for a given target language.⁴⁵ These results are related, but are not equivalent. For example, Tarskian indefinability does not imply inexpressibility—for a number of reasons. More on this shortly.

In the *Journal of Symbolic Logic* in 1950, John Myhill posed the question whether, given the negative result of Tarski's theorem, there were *any* logical systems that could express their own truth set. This challenge has spawned a series of results over the years demonstrating that various weakened or logically non-classical systems can be constructed which can have their own truth predicate.⁴⁶ Yet, the core acceptation of Tarski's result remains, albeit somewhat refined: *no logically classical system with at least the moderate expressive power of basic arithmetic can have its own truth predicate.*

set. If the given set is the truth set for that language, then the language "has its own truth predicate" (*albeit* not necessarily in the form of a simple predicate).

⁴⁴ Where exactly one is to find this result in Tarski's work is less than clear. One can find related results in Tarski (1935b; 1939); Tarski, Mostowski et al. (1953).

⁴⁵ Also, the former is thoroughly semantic and its proof presupposes things that the argument of the latter may not (Gómez-Torrente 2004).

⁴⁶ John Myhill (1950) showed that it could be done for a system weaker than PA which has no negation. Abraham Robinson (1963) showed that it could be done in a system with an infinitary logic. Saul Kripke (1975) showed it could be done with a three-valued logic. Anil Gupta (1982) showed it could be done classically if one put certain ad hoc restrictions on the extensions of terms.

Yet, while commonly accepted by philosophers, it is nonetheless not true, and the idea that Tarski's theorem establishes even such a moderate limit is spurious. Perforce, many will find this claim surprising, so I will briefly demonstrate the fact.⁴⁷

Let's take a simple characterization of the result, like this:

The language of arithmetic cannot have its own truth predicate.⁴⁸

A moment's reflection will tell you that such a result would be utterly trivial and takes no special cleverness to establish. The fact is that we know straight off that the language of arithmetic doesn't have the slightest chance of having its own truth predicate. We know this because the language of arithmetic does not have the sentences of the language of arithmetic in its domain of discourse. On its intended interpretation, the language of arithmetic has in its domain only numbers.⁴⁹ So, the above surely can't really be what we are learning about truth from this highly touted theorem. A more cautious formulation is this:

The language of arithmetic cannot have its own arithmetic truth predicate.

Typically this is what authors who use the earlier formulation implicitly or avowedly mean. However, *an arithmetic truth predicate is not a species of truth predicate* any more than a decoy duck is a species of duck. An arithmetic truth predicate (relative to some system of Gödel coding) would be a formula of the language of arithmetic which had, on the intended interpretation of that language, an extension consisting of the numeric codes of all and only the sentences of the language of arithmetic which are true (on the intended interpretation). So, while an arithmetic truth predicate does come by the "truth" in its name honestly enough, this does not make such a predicate a truth predicate. What that means for us is that neither of the above two statements of the result actually tells us anything straight-out about truth per se. They only sound like they are doing so. Instead they tell us something about the inexpressibility in the language of arithmetic of a set of numbers. Now, there is something interesting about that set of numbers, we shall grant, but if there is indeed some interesting, genuinely truth-relevant implication—call it a corollary—of this arithmetic result, neither of the above is a statement of it yet.

I elsewhere argue that philosophical interpretations of Tarski's result have been marred by something akin to a use-mention confusion which has masked the crucial role of coding and led many to an easy mis-statement/misunderstanding of the implications of this theorem. We will not go into this matter further here, but proceed

⁴⁷ I discuss this matter further in *Even Tempered Truth*, typescript.

⁴⁸ See e.g. Soames (1999: 148).

⁴⁹ Since the claim is about a truth predicate for the language, a fully interpreted language is presupposed, and what interpretation besides the intended interpretation of the language of arithmetic could be indicated here? None other unless one was being perverse.

directly to the surprising counterexample which proves the general understanding of Tarski's theorem is indeed wrong.

I offer a simple example of a system which has its own truth predicate (i.e. in which the truth set for that language is expressible). But unlike previous examples in the literature there will be nothing funny about the logic—it will be entirely classical. We will have a bivalent language with a wholly classical logic and semantics and we can outfit our system with as strong a set of arithmetic axioms as you like.⁵⁰

Let L be a standard first-order arithmetic language with the usual symbols ($+ \cdot \circ S < =$), and let A be the intended interpretation (i.e. an L -model with the natural numbers for domain and where “+” is interpreted as addition, etc.), and outfit our system with our favorite arithmetic theory, PA. Our logical system can thus be thought of as a triple, $\langle L, A, PA \rangle$.⁵¹

Now, there most certainly exists a set of the true sentences of this system and the axioms of PA are among them, and likewise for the set of falsehoods.

Now, let A^* be the interpretation for L obtained by replacing the even numbers in model A with the true sentences taken in some fixed order (e.g. orthographically ordered). Likewise, we put the false sentences in for the odd numbers. To make an isomorphic structure, we will also replace these items throughout the extensions assigned by the model A . Let ETS be just such a logical system, $\langle L, A^*, PA \rangle$, with language L and intended interpretation A^* .

Now, A^* is an L -model isomorphic to A , and so the very same sentences of L are made true as on our initial interpretation. It now follows immediately that this language has its own truth predicate and a very simple one at that:

there is a y such that $x = 2y$.

In short, the predicate *x is an even number* is a truth predicate for this new system in the sense that it expresses the truth set of the language. Thus, our language “has its own truth predicate” in the intended sense.

Thus we have an easy example of a bivalent system with a classical logic and outfitted with a consistent theory with the strength of PA which also unproblematically has its own truth predicate—which was to be proved.

Now, one would not want to attach too much significance to this example per se. Yet, if the common acceptance of Tarski's theorem were correct this example should not be possible. Its existence serves as a corrective, and reflection on it is useful in several ways. First, it can help us see what has gone missing in the understanding of Tarski's theorem. Second, it helps point up an essential distinction which must be attended to in order to sort out what is going on in this domain.

⁵⁰ Moreover, unlike Anil Gupta's denotation-restricted example, every element in the domain of discourse of our language will have a simple name.

⁵¹ The other aspects of a logical system (inference rules, rules of definition) are here entirely classical and pass without mention henceforward.

Let's take the second point first. Here is something to observe: ETS has its own truth predicate, and it also has a full set of T-sentences, i.e. for every sentence of the language, there is also a T-sentence for that sentence in the language.⁵² Yet, it is not hard to show that no Tarskian truth definition is formulable in this language: there is no formally correct definition which would be adequate in the sense of Convention T. The result is that the ETS system *has* its own truth predicate—it can express its own truth set and a truth predicate can be straightforwardly *defined*—and yet truth is Tarskian indefinable in this system. It is for this reason that we insisted before that one needs to carefully distinguish between a system *having a truth predicate* and its *having a Tarskian truth definition*.

I return now to the promised first point. How is it that the ETS system is able to have its own truth predicate without actually running afoul of Tarski's theorem? The answer is just this: the inexpressibility that Tarski's theorem would establish for an arithmetic language is *relative to a choice of coding function*. This coding function must have certain properties—suppressing details, let's just say it has to be an *effective* coding function. It appears that everyone who talks about Tarski's theorem simply assumes that, if the logical system in question has a certain minimum power—comes equipped with a strong enough theory to do something like very basic arithmetic—then an effective coding suited to getting the result is always available to be chosen.

In fact, they are almost right. An effective coding is always available, but one cannot always use it. For a *truth-relevant* result, not just any effective coding will do. It must be the *right* coding. If Tarski's theorem tells me in the first instance that the set of codes of true sentences is not expressible in my given language, then I do not yet know the *truth set* for that language itself is inexpressible. I could only infer that, if I could so choose my coding so that the set of codes of true sentences just was the set of true sentences—that is, by choosing the identity function (or permutation thereof) as my coding. Thus, Tarski's theorem only yields a genuinely truth-relevant result in a given case if there is available an effective, truth-relevant coding function. In the ETS system, there are truth-relevant codings and there are effective codings. There are many of them. But there are no effective truth-relevant codings.

Tarski's theorem identifies a real and definite limit on the expressibility of truth in logical systems. The foregoing does not cut against this, but it does, I hope, chasten us against drawing hasty conclusions about what that limit is.

- *Formal Indefinability.* We identified two other Tarskian limiting results: the formal indefinability theorem, and the essential richness thesis. The first of these is most closely identified with Theorem 1 of Tarski (1935b: 247), but might better be focused on Theses A and B from the 1935 postscript to that work (273), where Tarski enlarged on his fundamental result. In the latter parts of Tarski (1935b), Tarski addresses himself to the question of whether, given an object language, one could always find a suitable metalanguage in which to construct a Tarskian truth definition. Section 4 of that paper details

⁵² An intuitive most of these T-sentences will in fact be theorems of the system.

how this could always be done for object languages of so-called finite order. In Section 5, it is shown how the techniques of Section 4 would not be applicable to object languages of infinite order. Theorem 1 aims to establish that the barrier is a principled one. The theorem states that any proposed definition would make provable the negation of some T-sentence. This means that, if the object language system was theoretically consistent to begin with, a proposed definition that met Convention T could not be formally correct—since it would be a creative definition. Thus, Theorem 1 tells us there are no Tarskian truth definitions for languages of infinite order.⁵³

However, Tarski came to think that he had employed an unduly restrictive notion of language, and his reassessment of the central results of his paper are recorded in the 1935 postscript. Before, the highest order of a language was one which had variables of arbitrarily high finite order type. However, Tarski now saw no principled barrier to admitting variables of transfinite order type and hence of languages of transfinite order.

This new liberty enables Tarski to achieve a more general result and to cast in a new light what makes Tarskian definitions possible. Theses A and B distill this insight, stating (roughly) that, for a given object language, a metalanguage of higher order can always be found in which a Tarskian truth definition can be constructed, and that this cannot be done with a metalanguage which is not of higher order.

This result still embeds a limiting theorem, but the basic result from Tarski's point of view is all positive—in contrast to his entirely negative earlier result for languages of infinite order. And it is clear now why he had earlier gotten that negative result—his strategy of truth definition requires a metalanguage that in some sense stands above the object language in question, but the system of languages earlier allowed by Tarski “topped out” at languages of infinite order. Hence there could be no languages above (in the needed sense) an object language of infinite order. This observation is succinctly represented in Theses A and B.

• *Essential Richness Thesis.* Finally, we turn to the much maligned *essential richness thesis*. Tarski wrote that “the condition of the ‘essential richness’ of the metalanguage proves to be, not only necessary, but also sufficient for the construction of a satisfactory definition of truth” (Tarski 1944: 352). It is a frequent complaint about this claim that Tarski never tells us what makes one language essentially richer than another. David DeVidi and Graham Solomon (1999) have claimed that exhaustive analysis shows there is nothing Tarski *could* mean by the term to make the claim (non-trivially) true. I have argued elsewhere that these negative assessments are unfounded. Tarski tells us quite clearly in what the essential richness of a metalanguage consists, and the analysis of DeVidi and Solomon is quite fundamentally mistaken. Here, we will focus only on two points: 1) a clear statement of what would make a metalanguage be essentially richer than its object language, and 2) a clarified statement of the essential richness thesis.⁵⁴

⁵³ For Tarski's explanation of the notion of the order of a language see Tarski (1935b: 215–20). Cf. also Ray (2005).

⁵⁴ The interested reader is referred to Ray (2005) for the full story.

To say that a metalanguage *M* is *essentially richer than* an object language *L* is to say that one of the following obtains: i) *M* is a language based on the theory of semantic categories and *M* has variables of higher order than *L*; ii) the axioms of *M* prove the existence of something of higher order than can be proved to exist from the axioms of *L*; or iii) the rules of definition of *M* admit inductive definitions and the rules of definition of *L* do not.⁵⁵

Once we see what essential richness consists in, it is clear that Tarski's essential richness thesis is an informal statement of nothing other than a modified version of Theses A and B from Tarski (1935b). Rewriting those theses to take account of Tarski's twice-expanded notion of language enrichment, we get:

Essential Richness Thesis A. i) For semantically-open formalized language, *L*, and semantically open, exactly specified language, *M*, if *M* is essentially richer than *L* and if *M* is an extending metalanguage for *L*, then a formally correct and materially adequate semantic definition of true sentence of *L* can be constructed in *M* with the help only of general logical expressions, of the expressions of *L*, and of terms from the morphology of language. ii) For every such, *L*, there is at least one such *M* that is essentially richer than *L*, and which is an extending metalanguage for *L*.

Essential Richness Thesis B. For semantically open formalized language, *L*, and semantically open, exactly specified metalanguage, *M*, if a formally correct and materially adequate semantic definition of true sentence of *L* can be constructed in *M*, then *M* is essentially richer than *L*.

In 1944, Tarski claims straight out that being essentially richer is both necessary and sufficient for definability. The necessity is clear from ER Thesis B. The sufficiency is likewise clear from the italicized portion of ER Thesis A, given one important background assumption which is operative in Tarski's 1944 discussion, namely that *M* is an extending metalanguage for *L*.⁵⁶

26.7 CONCLUSION

Alfred Tarski was not trained as a philosopher in the sense that would be familiar today, and was rather more attuned to mathematical than philosophical matters, as is evidenced perhaps by his lifelong career in mathematics. The definitional concerns (Tarski 1931) which are precursors to his work on truth are as clearly mathematical in nature, as the work that follows upon it are logical (1939). So, perhaps it should not come

⁵⁵ Authority for this understanding is centrally rooted in Tarski (1935b: 271–2, 271 fn.1; 1944: 353).

⁵⁶ This requires that i) *L* is a sub-language of *M* and ii) *M* meets the conditions for being a metalanguage for *L*—and so e.g. has the ability to refer to the expressions of *L*.

as a surprise that, as we have found, Tarski's claims are often circumscribed in a way that does not admit of the sort of easy generality that might readily satisfy the philosopher in search of great themes. Key Tarskian ideas which have become mainstays of the philosophical discourse on truth—e.g. Convention T and T-sentences, the limiting theorems—are often not quite as simple or universally applicable as they are sometimes taken to be. On the question of Tarski's solution to the liar, we saw something similar.

Even something so seemingly central to our topic as *Tarski's theory of truth* has a way of not being quite what you would come to expect from philosophical sources. I once spent some time asking philosophers I met, "What is Tarski's theory of truth?" In answer, I was sure to get by way of reply a gesture toward some example Tarskian truth definition. Anecdotal evidence to be sure, but easily corroborated in the literature. The phrase "Tarski's theory of truth" appears most often to be used to "ambiguously refer" to any one of the infinitely many possible Tarskian truth definitions. One needs only inquire as to what Hartry Field's well-known essay, "Tarski's Theory of Truth" (1972), refers to by its title, or what answer Scott Soames has for the titular question of his essay, "What is a Theory of Truth?" (1984).

But surely it is clear that an arbitrary one of the indefinitely many possible Tarskian truth definitions cannot be Tarski's theory of truth—at least if we meant to earnestly ask what Tarski might have had to say that could answer to our philosophical interest in the concept of truth. Tarski does seem committed to a number of substantive views on the matter and even has arguments for some of them, but as we saw in section 26.4 it sometimes takes some teasing out. This is because, I think it is fair to say, Tarski saw his project as primarily logical, and only philosophical by its implications.

REFERENCES

- Badici, E. and K. Ludwig (2007). "The Concept of Truth and the Semantics of the Truth Predicate," *Inquiry* 50: 622–38.
- Chihara, C. (1979). "Semantic Paradoxes: A Diagnostic Investigation," *The Philosophical Review* 88: 590–618.
- Chihara, C. (1984). "Semantic Paradoxes: Some Second Thoughts," *Philosophical Studies* 45: 223–9.
- Davidson, D. (1990). "The Structure and Content of Truth," *Journal of Philosophy* 87: 279–328.
- DeVidi, D. and G. Solomon (1999). "Tarski on 'Essentially Richer' Metalanguages," *Journal of Philosophical Logic* 28: 1–28.
- Etchemendy, J. (1988). "Tarski on Truth and Logical Consequence," *Journal of Symbolic Logic* 53: 51–79.
- Field, H. (1972). "Tarski's Theory of Truth," *Journal of Philosophy* 69: 347–75.
- García-Carpintero, M. (1996). "What Is a Tarskian Definition of Truth?" *Philosophical Studies* 82: 113–44.
- Gómez-Torrente, M. (2004). "Indefinability of Truth in the *Wahrheitsbegriff*," *Annals of Pure and Applied Logic* 126: 27–37.
- Gupta, A. (1982). "Truth and Paradox," *Journal of Philosophical Logic* 11: 1–60. Reprinted in R. L. Martin (ed.) (1984), *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press, 175–235.

- Gupta, A. and N. Belnap (1993). *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Haack, S. (1978). *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Heck, R. (1997). "Tarski, Truth, and Semantics," *The Philosophical Review* 106: 533–54.
- Hodges, W. (2004). "What Languages Have Tarskian Truth Definitions?," *Annals of Pure and Applied Logic* 126: 93–113.
- Horwich, P. (1990). *Truth*, 1st edn. Oxford: Blackwell.
- Horwich, P. (1998). *Truth*. 2nd edn. Oxford: Oxford University Press.
- Kirkham, R. L. (1992). *Theories of Truth: A Critical Introduction*. Cambridge, MA: MIT Press.
- Kripke, S. (1975). "Outline of a Theory of Truth," *The Journal of Philosophy* 72: 690–716. Reprinted in R. L. Martin (ed.) (1984), *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press, 53–81.
- Luschei, E. C. (1962). *The Logical Systems of Leśniewski*. Amsterdam: North-Holland.
- Mates, B. (1972). *Elementary Logic*. 2nd edn. New York: Oxford University Press.
- McGee, V. (1992). "Maximal Consistent Sets of Instances of Tarski's Schema (T)," *Journal of Philosophical Logic* 21: 235–41.
- Myhill, J. (1950). "A System which can Define its own Truth," *Fundamenta Mathematicae* 37: 190–2.
- Patterson, D. (2012). *Alfred Tarski: Philosophy of Language and Logic*. New York: Palgrave Macmillan.
- Putnam, H. (1983). "On Truth," in Leigh S. Cauman et al. (eds.), *How Many Questions? Essays in Honor of Sidney Morgenbesser*. Indianapolis: Hackett, 35–56.
- Putnam, H. (1985). "A Comparison of Something with Something Else," *New Literary History* 17: 61–79.
- Ramsey, F. P. (1929) [1991]. *On Truth*. Episteme 16, ed. Nicholas Rescher and Ulrich Majer. Dordrecht: Kluwer.
- Ray, G. (2002). "Tarski, the Liar and Tarskian Truth Definitions," in D. Jacquette (ed.), *A Companion to Philosophical Logic*. Malden, MA: Blackwell, 164–76.
- Ray, G. (2003). "Tarski and the Metalinguistic Liar," *Philosophical Studies* 115: 55–80.
- Ray, G. (2005). "On the Matter of Essential Richness," *Journal of Philosophical Logic* 34: 433–57.
- Ray, G. (2006). "Tarski's Grelling and the T-Strategy," in B. Brown (ed.), *Truth and Probability: Essays in Honour of Hugues Leblanc*. London: College Publications.
- Ray, G. (2010). *Even Tempered Truth*. Typescript.
- Ray, G. and K. Ludwig (2002). "Vagueness and the Sorites Paradox," in J. Tomberlin (ed.), *Language and Mind*. Atascadero, CA: Ridgeview Press, 419–61.
- Robinson, A. (1963). "On Languages which are Based on Nonstandard Arithmetic," *Nagoya Mathematical Journal* 22: 83–117.
- Russell, B. (1912). "Truth and Falsehood," in B. Russell, *Problems of Philosophy*. New York: Henry Holt, 196–203.
- Soames, S. (1984). "What is a Theory of Truth?" *The Journal of Philosophy* 81: 411–29.
- Soames, S. (1995). "T-Sentences," in W. Sinnott-Armstrong, D. Raffman, and N. Asher (eds.), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press, 250–70.
- Soames, S. (1999). *Understanding Truth*. New York: Oxford University Press.
- Tarski, A. (1931). "Sur les ensembles définissables de nombres réels (I)," *Fundamenta Mathematicae* 17: 210–39. Trans. J. H. Woodger as "On Definable Sets of Real Numbers," in Tarski, *Logic, Semantics, Metamathematics*, 2nd edn. (1983). Indianapolis: Hackett: 110–42.

- Tarski, A. (1933). *Pojęcie prawdy w językach nauk dedukcyjnych*. Warsaw: Nakładem Towarzystwa Naukowego Warszawskiego.
- Tarski, A. (1935a). "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica* 1: 261–405.
- Tarski, A. (1935b). "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica* 1: 261–405. Trans. J. H. Woodger as "On the Concept of Truth in Formalized Languages," in Tarski, *Logic, Semantics, Metamathematics*, 2nd edn. (1983). Indianapolis: Hackett: 152–278.
- Tarski, A. (1936a). *O logice matematycznej i metodzie dedukcyjnej*. Lwów and Warsaw: Książnica-Atlas. Trans. O. Helmer and revised from the 1937 German translation of the Polish as Tarski, *Introduction to Logic and to the Methodology of Deductive Sciences* (1941). Oxford: Oxford University Press.
- Tarski, A. (1936b). "O pojęciu wynikania logicznego," *Przegląd Filozoficzny* 39: 58–68. Trans. J. H. Woodger as "On the Concept of Logical Consequence," in Tarski, *Logic, Semantics, Metamathematics*, 2nd edn (1983). Indianapolis: Hackett: 409–20.
- Tarski, A. (1936c). "O ugruntowaniu naukowej semantyki," *Przegląd Filozoficzny* 39: 50–7. Trans. J. H. Woodger as "The Establishment of Scientific Semantics," in Tarski, *Logic, Semantics, Metamathematics*, 2nd edn (1983) Indianapolis: Hackett: 401–8.
- Tarski, A. (1939). "On Undecidable Statements in Enlarged Systems of Logic and the Concept of Truth," *Journal of Symbolic Logic* 4: 105–12.
- Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341–76.
- Tarski, A. (1969). "Truth and Proof," *Scientific American* 220: 63–77.
- Tarski, A., A. Mostowski, and R. Robinson (1953). "Undecidability and Essential Undecidability in Arithmetic," in A. Tarski, A. Mostowski, and R. Robinson (eds.), *Undecidable Theories*. Amsterdam: North-Holland, 40–74.

CHAPTER 27

THE AXIOMATIC APPROACH TO TRUTH¹

KENTARO FUJIMOTO AND VOLKER HALBACH

27.1 CONCEPTUAL ANALYSIS AND DEFINITIONS

TRADITIONALLY, the conceptual analysis of a notion often takes the form of an explicit definition. Philosophers have proposed explicit definitions of acting freely, knowing, personhood, and other notions. Of course, in each case they have had disagreements about which definition is adequate, but with respect to many notions they have shared an optimism about the possibility of a nontrivial adequate definition. For instance, the assumption that a definition of knowledge is possible seems to have been presupposed in many discussions throughout the history of epistemology.

Over time, however, doubts have cropped up about the possibility of an informative adequate definition of the notions that have traditionally been analyzed by means of explicit definitions. In the case of knowledge, Gettier problems and numerous failures in solving them have nourished the suspicion that there are no nontrivial definitions of knowledge in terms of truth, justification, and belief, and perhaps no illuminating definition of knowledge at all. Nowadays many philosophers seem happy to take knowledge as a primitive undefined notion.²

Also in the case of truth, philosophers have proposed explicit definitions. They have disagreed about whether truth can and should be defined in terms of correspondence, coherence, utility, or in physicalistic terms; but the presupposition that there *is* an informative and nontrivial definition has often gone unchallenged. Now more philosophers

¹ This work was supported by the Arts & Humanities Research Council AH/H039791/1.

² See, for instance, Williamson (2000).

have become prepared to give up on an explicit definition of truth and to take truth as a primitive notion.

Compared to the case of knowledge, there is further evidence beyond the failure of many attempted definitions for the view that truth should be taken as primitive: Tarski (1935) proved that a definition of truth is impossible under fairly general assumptions.³

In order to obtain a conceptual analysis without presupposing that an interesting definition of truth is feasible, one can study the principles guiding our reasoning with the truth predicate. These principles can be made explicit by spelling them out as axioms and rules for truth. This is the axiomatic approach to truth.

Axiomatic approaches to truth, knowledge, or other notions do not rule out the possibility of a definition for the notion in question: a notion that is taken as primitive can turn out to be definable after all. On an axiomatic approach, one just does not presuppose that the notion is definable. Once it has been set out what is expected from the notion, one can investigate whether a notion satisfying these conditions is definable or eliminable in some sense.

27.2 FORMAL AXIOMATIC THEORIES OF TRUTH

Very often, when talking about axiomatic theories of truth, logicians and philosophers refer to *formal* axiomatic theories; and here we will also survey the formal approaches.

A formal approach facilitates the analysis of the axioms and rules for truth. The tools of formal logic can be used to investigate whether a truth predicate satisfying certain axioms and rules is consistent or definable in a certain theory or whether it is reducible to a system by some means. One can also investigate whether a system of truth yields any new consequences outside the theory of truth. The question of whether certain axioms for truth have such consequences is highly relevant for the discussion about deflationary conceptions of truth.⁴

On these approaches truth is conceived as a predicate; and the truth predicate applies to objects—whether they are propositions, types or tokens of sentences, or still something else. So, the axioms for truth are usually adjoined to a theory that codifies assumptions about the objects to which truth is attributed.

We call this theory the *base theory*. Beyond the theory of these objects, the base theory may contain further assumptions on other kinds of objects. Usually, the objects of truth are taken to be the sentences or corresponding propositions of the language of the base theory, possibly enlarged with the truth predicate.

³ See Ray (ch. 26 in this volume) on Tarski's theory.

⁴ See Azzouni (ch. 17 in this volume).

Without a base theory, many axioms for truth could not be sensibly formulated. For instance, an axiom expressing that a conjunction is true if and only if both conjuncts are true only makes sense on the background of certain assumptions about what a conjunction is. Thus, for instance, the base theory should prove that a conjunction of sentences (or propositions) is different from any negated sentence (proposition).

On most formal approaches, truth is attributed to sentences or codes of sentences. The theory about these objects can often be interpreted as a theory of propositions, as long as propositions are assumed to be structured like sentences, as long, that is, as there are atomic propositions expressed by atomic sentences of the chosen language, and there are operations on propositions that work like conjunction, negation, etc. on sentences.

In order to avoid the need for developing a theory of expressions or propositions from scratch, many logicians avail themselves to Gödel's technique of theorizing about expressions in arithmetic via coding and use first-order Peano arithmetic as base theory. Here we will also employ Peano arithmetic as base theory although most of the results stated in the present chapter still obtain if some other base theory is used. Large parts of the discussion below apply with small changes to alternative base theories as well. At the end we will give some hints on how to apply axiomatic theories of truth to other base theories.

Gödel showed that the language of arithmetic can be coded by the numbers: each expression, that is, each finite string of symbols in that language, is assigned a number in some systematic way. Since the language of first-order arithmetic contains names (numerals) for numbers, one can think of the language of arithmetic as containing names for all its expressions (or rather their codes; but we suppress this distinction in what follows). We will write $\ulcorner \phi \urcorner$ for the numeral of (the code of) an expression ϕ . These numerals for sentences will play the role of "structural-descriptive" names of sentences in Tarski's (1935) sense.

As just outlined, any theory of truth we will consider is an extension of the base theory, that is, Peano arithmetic. The main question we'll consider is, which truth axioms should be added to this base theory? We can take inspiration from Tarski. In his celebrated Convention T (1935: 187f) Tarski stipulates that a definition of truth is adequate if and only if the T-biconditionals for the object-language and the claim that only sentences of the object-language are true are derivable from the metatheory and the definition.⁵

On an axiomatic approach, we can simply add a truth predicate with those properties. That is, we add a new unary predicate T to the language \mathcal{L} of arithmetic. This yields the language \mathcal{L}_T . Then we augment Peano arithmetic with all sentences

$$T\ulcorner \phi \urcorner \text{ iff } \phi$$

⁵ This exposition is highly simplified. For a more detailed discussion, see Ray (ch. 26 in this volume) and Halbach (2011).

where ϕ is a sentence of the language of arithmetic. The system $TB\downarrow$ is formulated in classical logic, and so will be all systems below unless stated otherwise.⁶ We call the resulting theory $TB\downarrow$ (TB stands for *Tarski biconditionals*; the purpose of the symbol \downarrow will be explained later on). To keep the system simple, we do not add an axiom stipulating, as in the second condition of Convention T, that only sentences of the language of arithmetic are true; adding such an axiom would not make any great difference.

At this point we would like to point to an incoherence in Tarski's stance on $TB\downarrow$, as his rejection of $TB\downarrow$ is in conflict with his other views. Tarski (1935: 257) observed that the *principle of contradiction*, that is, the universal claim that no sentence (without the truth predicate) can be true together with its negation, and similar general claims are not derivable in the theory $TB\downarrow$ (although one can prove of any given single sentence of this language that it cannot be true together with its negation).⁷

Of course the principle of contradiction can be added to $TB\downarrow$ as a new truth axiom, but Tarski "attach[es] little importance to this procedure" as "it seems that every such enlargement of the axiom system has an accidental character" (1935: 258).

If Tarski is justified in his rejection of $TB\downarrow$, then his adequacy condition in Convention T is too weak. In Convention T he calls any definition of truth adequate as long as the T -biconditionals can be proved for the object-language, that is, the language of arithmetic here. In the theory $TB\downarrow$ a truth predicate Tr that is adequate by Tarski's own standards can trivially be defined: a definition of a truth predicate Tr as T trivially satisfies Convention T.⁸

But the principle of contradiction fails to be provable for Tr as well, because $TB\downarrow$ does not prove the principle for T and Tr is defined as T . Now, if the principle of contradiction should be provable for any decent truth predicate for the language of arithmetic, then Tr fails to be a decent truth predicate. There are of course alternative definitions of Tr in richer metalanguages that are adequate by Tarski's standards, such that the principle of contradiction becomes provable for Tr . So it seems that Tarski would have to concede that among the adequate definitions there are definitions that are better or "more adequate" than others. This is a serious problem for Tarski's account: if he rejects $TB\downarrow$ as too weak, he is committed to rejecting definitions of truth as too weak that are adequate by the standards of his Convention T.

Tarski might reject the definition of the truth predicate Tr in the disquotational theory $TB\downarrow$ by insisting that all adequate definitions of truth must be in purely non-semantic terms. On this account, the definition of Tr as the truth predicate T of $TB\downarrow$ would not qualify as an adequate definition, because Tr is defined in terms of the semantic predicate T . So one might cherish the hope that all adequate definitions of truth

⁶ Non-classical approaches will be mentioned briefly at the end of this chapter.

⁷ Tarski proved this claim for a different framework. We have adapted Tarski's results to the set-up used here.

⁸ Here we are suppressing again the condition in Convention T that stipulates that only sentences of the object-languages are true. A truth predicate $Tr(x)$ satisfying this condition can be defined as $Tx \wedge \text{Sent}(x)$ where $\text{Sent}(x)$ expresses that x is a sentence of the object-language, that is here, of the language \mathcal{L} of arithmetic.

in purely non-semantic terms will yield the principle of contradiction and similar generalizations. However, Mostowski (1950) provided a natural example of a truth definition in purely non-semantic terms that is adequate by Tarski's standards but does not imply general claims such as the principle of contradiction.

After all, it seems that the adequacy condition of Convention T is too weak by Tarski's own standards. The T-biconditionals must be derivable in any adequate theory of truth for the language of arithmetic, but other principles should be provable as well and Tarski will have to embark on the project of finding suitable enlargements of theories like TB| by suitable principles whether they are used as new axioms, as we would prefer, or whether they are used as further adequacy conditions, as Tarski should probably prefer.

In fact Tarski's approach can be seen as axiomatic: first he states axioms for truth (mainly the T-biconditionals)—and then he goes on to show that this truth predicate can be defined in an "essentially richer metalanguage," or, more precisely, in a richer metatheory.⁹

However, one may not want or be able to go on to a richer metatheory containing further non-semantic assumptions. In particular, if we assume that our present state of knowledge can be captured by a formal system, then there is no way to define the truth predicate for this theory or language without moving on to a stronger theory, simply because no such stronger theory is available to us. By introducing new commitments and assumptions we may be able to define the truth predicate, but one may prefer to avoid new commitments and assumptions and to introduce the truth predicate via axioms and thereby evade the need for additional non-semantic commitments.

We suspect that Tarski was very much pursuing a reductive program: he probably preferred to introduce new mathematical assumptions (about the existence of sets, for instance), rather than making irreducible semantic assumptions. Many philosophers have followed Tarski in this approach. In particular, in truth theories like Kripke's (1975) *Outline of a Theory of Truth* or in the revision theory of Belnap and Gupta (1993), and others, a semantics for a language with a truth predicate is defined in set theory.

Kripke's approach and many other "semantic" theories of truth share one central feature with Tarski's theory: all semantic notions are eliminated by definitions in terms of mathematical notions and thereby reduced to mathematics, usually to set theory. But it is not possible to apply the semantic approaches to the languages and theories one is working in; these theories require a stronger metalanguage in order to carry out the semantic constructions.

The axiomatic approach is superior in this respect: if set-theoretic reductionism is abandoned and truth is accepted as a primitive undefined notion, axioms and rules for

⁹ Tarski tends to use the term *language* also for formal systems or theories. So where he uses the term *metalanguage* it may be more apt to use *metatheory* in modern terminology. The discussion of Tarski's approach here is very sketchy, of course. See, for instance, Heck (1997).

truth can be added to any base theory without the need to make further non-semantic commitments. In particular, (Zermelo-Fraenkel) set theory can be used as base theory without the need for further mathematical axioms that would be needed for carrying out a semantic definition of truth. Of course in the axiomatic approach, one needs to talk about axiomatizations and formal theories, but theories far weaker than set theory or Peano arithmetic are already sufficient for reasoning about these syntactic notions.

As we shall see below, however, the semantic and the axiomatic approach are closely linked, and many axiomatic theories have taken their inspiration from semantic approaches.

27.3 THEORIES OF TYPED TRUTH

In order to overcome the problem of the unprovability of the law of contradiction, Tarski seemed to lean toward a solution that involves the strengthening of the underlying logic by an infinitary rule of inference, namely a rule that has come to be known as “ ω -rule.” However, by adopting this rule, one is leaving the realm of formal systems with a decidable notion of proof; Tarski’s proposal has met with little approval.

Logicians and philosophers went down the route that Tarski rejected as unpromising: They added further general principles beyond the T-biconditionals as axioms for truth. Tarski’s own work prepared the way for this. Most of the general laws such as the law of contradiction that are expected to be theorems of an adequate theory of truth follow from the clauses of an inductive “Tarskian” definition of truth or satisfaction. Thus, if the clauses of Tarski’s definition of truth or satisfaction are turned into axioms and added to a reasonable base theory of syntax, chances are high that the resulting theory proves all general principles that ought to be provable in an adequate truth theory.

In various papers collected in his *Inquiries into Truth and Interpretation* (1984), Davidson made the axiomatic counterpart of Tarski’s theory of truth popular among philosophers of language, without ever specifying precisely the axioms of that theory even for a simple base theory in a precise way.

Focusing on the present setting with Peano arithmetic as the base theory, truth could be axiomatized in \mathcal{L}_T by the formalizations of the following claims in addition to the PA-axioms, where *sentence* stands for *sentence in the language \mathcal{L} of arithmetic (without T)*:

- C1 A sentence of the form $s = t$ (with s and t closed terms) is true iff s and t coincide in their values. (The latter condition can be expressed in PA.)
- C2 A sentence of the form $\neg\phi$ is true iff the sentence ϕ itself is not true.
- C3 A sentence of the form $\phi \wedge \psi$ is true iff ϕ is true and ψ is true.
- C4 A universally quantified sentence $\forall x\phi(x)$ is true iff all its substitution instances $\phi(t)$ are true. (Here $\phi(t)$ stands for the result of substituting all free occurrences of the variable x with the closed term t).

The result of adding these axioms to PA yields the theory $CT\downarrow$. The *principle of contradiction* is a logical consequence of the second axiom, so it is a consequence of that theory. Also all T-biconditionals for sentences without T are provable in $CT\downarrow$; hence $TB\downarrow$ is a proper subtheory of $CT\downarrow$.

The acronym $CT\downarrow$ stands for *compositional truth*, as the axioms $c2$ – $c4$ are compositional in the sense that they ascribe a semantic status—*truth* or *falsity*—to a sentence dependent on the semantic status of the direct subsentences. Since we are working within classical logic, $c2$ implies that the truth conception of $CT\downarrow$ is bivalent: for each sentence σ , either σ itself or its negation $\neg\sigma$ but not both is true. Then, conceiving the falsity of a sentence σ as the truth of its negation $\neg\sigma$, the compositional axioms for falsity are all derivable from $c1$ – $c4$; for instance, it follows from $c2$ and $c3$ that a conjunction is false if at least one of the conjuncts is false. Accordingly, $CT\downarrow$ describes classical bivalent semantics in which every sentence takes a single truth value (*truth* or *falsity*) and truth values are assigned to complex sentences according to the rules of classical logic.

If the base language contains predicate symbols beyond the identity symbols, an axiom analogous to $c1$ should be added for each further predicate symbol; and if there are connectives and quantifiers in the language beyond negation, conjunction, and the universal, quantifier axioms analogous to $c2$ – $c4$ must be added. In what follows, however, we assume that the base language \mathcal{L} contains the identity symbol as its only predicate symbols and \neg , \wedge , and \forall as its only logical symbols.

The system $CT\downarrow$ is conservative over PA for arithmetical sentences in the sense that if a sentence of the language of arithmetic is provable in $CT\downarrow$ it is provable already in PA without any truth-theoretic axioms.¹⁰ But not any model of PA can be expanded to a model of $CT\downarrow$, as was shown by Lachlan (1981).

The theory $CT\downarrow$ contains all induction axioms, that is, all axioms of the form

$$\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(x+1)) \rightarrow \forall x\phi(x)$$

for $\phi(x)$ without the truth predicate T , as they are already theorems of PA, but not all induction axioms for formulae *with* the truth predicate are derivable in $CT\downarrow$. The absence of induction axioms seems to be an unnatural restriction. In particular, many natural arguments on the length of formulae or proofs cannot be carried out in $CT\downarrow$, because the required instances of the induction schema are not available in $CT\downarrow$.¹¹

In what follows the symbol \downarrow at the end of a name for a system will indicate that only instances of the induction schema without the truth predicate are axioms of the system.

¹⁰ A model-theoretic result that implies the conservativity was proved by Kotlarski et al. (1981). Building on an incorrect proof by Halbach (1999), Leigh (2013) proved conservativity by proof-theoretic methods. Enayat and Visser (2015) presented a more elegant model-theoretic proof.

¹¹ If a theory of expressions or propositions without the detour via coding in arithmetic were used as base theory, an axiom schema for induction on the grammatical structure of expressions or propositions could be used. Their treatment would correspond in many details to the axiom schema for induction on natural numbers.

Without \uparrow all induction axioms are assumed. Hence adding all induction axioms in the language *with* the truth predicate to $\text{CT} \uparrow$ yields the theory CT .

By an induction on the length of proofs (and using an instance of induction with T for proving that the universal closures of all induction axioms of arithmetic are true), it is provable in CT that all closed theorems of PA are true. In particular, it can be shown in CT that if $0 = 1$ were provable in PA , $0 = 1$ would be true. But since the refutable claim $0 = 1$ is equivalent to $T \ulcorner 0 = 1 \urcorner$ by Axiom 1, the contradiction $0 = 1$ is not provable in PA , and therefore PA must be consistent. Since the consistency of PA is provable in CT but not in PA by Gödel's second incompleteness theorem, the theory CT is not conservative over PA .

This observation is a blow to those claiming that truth cannot serve any “explanatory” purpose or that truth cannot yield any new substantial insights. This may be a problem for deflationism about truth. See Halbach (2011) for an overview.

27.4 TYPE-FREE TRUTH

The typed truth theories in the preceding sections still rely on Tarski's solution of the liar paradox. They respect the distinction of object- and metalanguages in its axioms in the following sense: the truth predicate T belongs to the entire language \mathcal{L}_T , but sentences containing T are all excluded from the intended scope of the predicate T , and T can be sensibly applied only to a proper fragment \mathcal{L} of the entire language \mathcal{L}_T . However, Tarski's method is neither the only way, nor the best way for every purpose—to resolve the paradoxes. Many philosophers reject this typed conception of truth as unsuitable for their aims, in particular for the analysis of the notion of truth in our natural language. How can we read this object-/meta distinction into our natural language? Our natural language contains the predicate “is true,” but it apparently embodies no type classification.

Imposing the object-/metalanguage distinction on natural language excludes many natural expressions from our natural language as “invalid” or “ill-formed,” as many authors, in particular Kripke (1975), have argued. For instance, suppose Adam says to Ben, “You are very honest. You only tell the truth to me,” and Ben replies to Adam, “Thank you. You do as well.” In other words, they say to each other that “whatever you say to me is true.” Then, at least after the utterance of Ben, we cannot determine to which proper sublanguage of their natural language the predicate “is true” used by Adam or Ben belongs. Consequently, one would argue that a typed conception of truth fails to capture some natural and essential aspects of the notion of truth in our natural language.

Because of this discontent with typed truth, a variety of approaches for a theory of type-free truth have been attempted and emphasis has thus shifted from typed to type-free truth in the formal study of truth. For this purpose, we need to consider a theory of

truth in which the truth predicate can be properly applied to sentences which may contain the truth predicate itself.

However, no language can contain a predicate for type-free truth in the unrestricted and most naïve sense. The liar paradox tells us that under fairly general conditions we can't have a truth predicate T which validates the full T-schema: that is, the schema of all the T-biconditionals for arbitrary sentences possibly containing T itself.

In the study of type-free truth, we are thus always exposed to the danger of contradiction (because not only of the liar paradox but also of further paradoxes). Once the Tarskian distinction between object- and metalanguages is abandoned, great caution must be exercised to avoid inconsistency. Furthermore, inconsistency is not the only problem threatening type-free theories of truth. A consistent system with apparently natural axioms can suffer from drawbacks nearly as bad as inconsistency. For instance, there are truth theories proving unsound statements of the base language.¹² Other theories have only trivial models: for instance, only models with an empty extension of the truth predicate. Still other theories are ω -inconsistent.¹³ These facts urge us to somehow refine the naïve conception of truth, and restrict some natural principles of truth, in order to escape from inconsistency and properly deal with the other aforementioned problems.

In spite of these problems, philosophers and logicians have made considerable effort to provide adequate theories of type-free truth and numerous proposals have hitherto been presented. In what follows, we will introduce and examine some of the major axiomatic theories of type-free truth, each of which bears its particular conception of truth and has its advantages with respect to others.

27.4.1 Disquotational theories of type-free truth

The liar paradox shows that no theory with a sufficiently strong base theory can contain all T-biconditionals

$$(4.1) \quad T \ulcorner \sigma \urcorner \text{ iff } \sigma,$$

for every sentence σ , whether or not it contains the predicate T itself. Instantiating the schema with a liar sentence, that is, a self-referential sentence λ such that $\neg T \ulcorner \lambda \urcorner$ iff λ immediately yields a contradiction together with (4.1) for λ . However, we can obtain systems of type-free truth by somehow restricting the full T-schema to a proper subschema. We call theories of truth based on such a subschema *disquotational*.

¹² By “unsound statements” we mean those statements which are false in the intended (standard) model of the base theory; e.g. in our setting where the base theory is arithmetical, such statements correspond to false arithmetical statements in the standard model \mathbb{N} of arithmetic.

¹³ A theory is ω -inconsistent iff it proves $\phi(\bar{n})$ for every numeral \bar{n} ($n \in \mathbb{N}$) but also $\exists x \neg \phi(x)$.

We have observed that at least the above λ ought to be excluded from any consistent subschema of the full T-schema. However, we cannot block inconsistency only by excluding λ , as there are many other T-biconditionals which yield inconsistency. In general, it is not at all easy to sieve out consistent subschemata of the full T-schema. Of course, the obvious case in point is the subschema obtained by restricting σ in (4.1) to the sentences containing no occurrence of T ; this results in the theory $TB\upharpoonright$ but it is a theory of typed truth. Since any disquotational theory is expected to include $TB\upharpoonright$, the disquotational approach is understood as an attempt at extending the disquotational theory $TB\upharpoonright$ for typed truth to some adequate disquotational theory for type-free truth.

It is sometimes thought that disquotational theories should be fairly weak, probably because typed disquotational theories such as $TB\upharpoonright$ and TB are conservative over PA. However, the T-biconditionals are in fact a very strong device, not in the trivial sense that certain sets of T-biconditionals are inconsistent and derive everything, but in the sense that even consistent sets of T-biconditionals can have unlimited and arbitrary deductive strength.

It follows from McGee's trick in McGee (1992) that over Peano arithmetic, *any* sentence is equivalent to a certain T-biconditional. More precisely, given any sentence ϕ , we can construct, by using Gödel's technique of diagonalization, a sentence ψ containing T so that the following equivalence is provable in Peano arithmetic:

$$\psi \leftrightarrow (T\ulcorner \psi \urcorner \leftrightarrow \phi)$$

By the associativity of the biconditional in classical logic, the T-biconditional for this ψ , that is $\psi \leftrightarrow T\ulcorner \psi \urcorner$, is equivalent to ϕ . This lambent observation by McGee, which is reminiscent of Curry's paradox, should be a decisive counter-argument to the view that disquotational (or minimalist or identity) theories of truth are bound to be weak; any set of sentences, in particular, any set of truth axioms, can be reaxiomatized over a weak theory of arithmetic as a set of T-biconditionals.

So it turns out that type-free disquotational theories can have arbitrary strength, in contrast to their typed analogue. But how then should we select a consistent subschema and what kind of subschema is to be chosen from this vast range of options? Unless one sticks to Tarski's type distinction, the main challenge for the disquotationalist is the problem of specifying a sensible set of T-biconditionals.

Some proposals have already been made. For example, one might want not to exclude any disquotation sentence unless necessary and to endorse as many disquotation sentences as is consistently possible. Horwich (1990) adopts this approach (with the difference that truth is attributed to propositions rather than sentences). He expresses his hope of arriving at an attractive theory of truth by adopting a maximal consistent subset of the full T-schema.

However, McGee (1992) showed that there are uncountably many mutually inconsistent maximal consistent subsets of the set of all T-sentences, that is, the full T-schema. In fact, McGee's result tells us more: for each consistent subset \mathcal{S} of the full T-schema, we have a maximal consistent subset including \mathcal{S} . This observation indicates that

Horwich's proposal cannot guide us to decide between the maximal consistent subsets of the full T-schema, because there are uncountably many such maximal consistent sets. Moreover, as McGee shows, a maximal consistent subset will decide any sentence; and by choosing a particular set, any independent sentence can be made provable or refutable. This fact shows that inconsistency is not the only possible threat from paradox: if the theory tells us who will be the next US president, something is wrong with our truth theory.

Faced with this failure, one might try to strengthen the maximal consistency principle by adding further conditions; for example maximal ω -consistency, maximal soundness, and maximal conservativity (over the base theory PA); however, it is already known that none of these suffices for the purpose.¹⁴

Instead of maximality one can try to specify more syntactic restrictions. Halbach (2009) proposes considering T-positive instances. Roughly speaking, these are instances in which no occurrence of the truth predicate is in the scope of a negation symbol. If moreover the T-biconditionals are stated in a parametrized, uniform way—for instance in the following form:

For all closed terms t : ϕ is true of the value of t iff $\phi(t)$,

where $\phi(x)$ is any T-positive formula with only x free—then the resulting system PUTB, which includes all induction axioms in the language \mathcal{L}_T , is very strong. In fact it is as strong as the powerful type-free theory KF, which will be discussed below.

27.4.2 Classical and symmetric theories of truth

In the last subsection we considered extending $TB \upharpoonright$ or TB (with all induction axioms) to theories of type-free truth. Now we generalize the compositional theories $CT \upharpoonright$ or CT to a system of type-free truth. This is easily achieved by quantifying over all sentences in the extended language with truth in the truth-theoretic axioms C1–C4. So the axioms are exactly the same as C1–C4 with the exception that *sentence* is now understood as *sentence in the language \mathcal{L}_T* rather than \mathcal{L}_{PA} :

- F1 For arbitrary closed terms s and t , a sentence $s = t$ is true iff s and t coincide in their values.
- F2 An \mathcal{L}_T -sentence of the form $\neg\phi$ is true iff the sentence ϕ itself is not true.
- F3 An \mathcal{L}_T -sentence of the form $\phi \wedge \psi$ is true iff ϕ is true and ψ is true.

¹⁴ McGee (1992) established that, as in the case of maximal consistency, there are uncountably many mutually inconsistent subschemata that are maximal consistent by ω -logic. Cieřliński (2007) later showed that there are also uncountably many mutually inconsistent maximal conservative subschemata.

- F4 A universally quantified \mathcal{L}_T -sentence $\forall x\phi(x)$ is true iff all its substitution instances $\phi(t)$ are true.

By a parallel argument to the derivability of the principle of contradiction for \mathcal{L} -sentences in CT, the same principle even for \mathcal{L}_T -sentences is derivable from the axiom F2. In fact, we can derive many other general principles from these axioms F1–F4. In particular, the principle of excluded middle, “every sentence is either true or false,” is derivable. Therefore, the truth conception represented by F1–F4 is bivalent. Then, by equating the falsity of a sentence with the truth of its negation, the compositional axioms for falsity are all derivable from F1–F4 as in the case of CT \upharpoonright . Hence these axioms describe a classical conception of truth.

The axioms F1–F4 do not stipulate anything regarding the truth or falsity of atomic sentences with the truth predicate. Since we are pursuing a theory for type-free truth, which is supposed to have no distinction of object- and metalanguages, we naturally want to treat the truth predicate T in the same manner as the other predicates of the base language. Hence, it seems plausible to consider adding the following axiom together with F1:

- F5 For an arbitrary closed term t , the sentence Tt is true iff the value of t is a true \mathcal{L}_T -sentence.

This axiom can be also regarded as expressing the iterability of truth in the sense that the truth can be repeatedly attached to or detached from the truth of a sentence arbitrarily many times. Unfortunately, however, we can derive a contradiction from F1–F5. In fact, we can show by induction on the complexity of formulae that F1–F5 derive the full T-schema (4.1) and are therefore inconsistent over Peano arithmetic.

As in the case of the disquotational approach, we accordingly need to restrict the full compositionality of type-free truth represented by F1–F5. One obvious way is to simply drop F5. Let $FS_0\upharpoonright$ be the theory PA plus F1–F4; it is easily seen that CT \upharpoonright is a subtheory of $FS_0\upharpoonright$. Although the converse fails (and thus CT \upharpoonright is a *proper* subtheory of $FS_0\upharpoonright$), we can show that $FS_0\upharpoonright$ is conservative over CT \upharpoonright for arithmetical sentences; therefore, $FS_0\upharpoonright$ is as safe as CT \upharpoonright and PA from the viewpoint of consistency. Moreover, adding all induction axioms for \mathcal{L}_T does not add much strength. Following our stipulation in 27.3, the resulting theory is written as FS_0 . Then, we can also show that CT is a proper subtheory of FS_0 and that FS_0 is conservative over CT for arithmetical sentences.

Having seen the strong safety from inconsistency as well as the aforementioned derivability of general principles, $FS_0\upharpoonright$ or FS_0 may appear to be a satisfactory theory of type-free truth at a first glance. However, without the help of F5, neither $FS_0\upharpoonright$ nor FS_0 can derive the truth of any sentence containing the truth predicate T ; neither can derive $T^\top T^\top 0 = 0^\top$, for example. This seems a crucial defect of $FS_0\upharpoonright$ and FS_0 as a theory of type-free truth.

Halbach (1994) addressed this problem by adding two rules of inference to FS_0 instead of the full iterability axiom F_5 :

- From ϕ we can infer $T'\bar{\phi}^\top$ (for any sentence ϕ with or without the truth predicate).
- From $T'\bar{\phi}^\top$ we can infer ϕ (for any sentence ϕ with or without the truth predicate).

The resulting theory is named FS, after Friedman and Sheard (1987).¹⁵ Friedman and Sheard provided a model-theoretic construction through which the consistency of FS can be established. Halbach (1994) later showed that FS is as strong as the iteration of CT through all finite levels: an axiomatization of the hierarchy of “Tarskian” truths with the predicates $T_0, T_1, T_2, T_3, \dots$, where T_n is the compositional truth in the sense of 27.3 for the language only containing T_0, \dots, T_{n-1} .

It immediately follows from the definition of FS that $FS \vdash \psi$ and $FS \vdash T'\bar{\psi}^\top$ are equivalent. This implies coincidence between the provable sentences and the provably true sentences within FS. We call the set $\{\psi \mid FS \vdash T'\bar{\psi}^\top\}$ the *inner theory* of FS, while the set $\{\psi \mid FS \vdash \psi\}$ we call the *outer theory* of FS by contrast. Thus, the inner theory and outer theory coincide in the case of FS; this indicates that the inner theory of FS follows classical logic as well as its outer theory (i.e. FS itself). This desirable property, the coincidence between inner and outer logics, is counted by Leitgeb (2007) among his list of desiderata for adequate theories of truth. By contrast to FS, many other theories of type-free truth fail to possess this property. In this respect, FS is a very well-behaved theory.

However, FS still has some serious defects. First of all, a result by McGee (1985) implies that FS is ω -inconsistent, although it is arithmetically sound; see Halbach (1994). This implies that FS has no standard model and thus fails to meet another desideratum in the aforementioned list by Leitgeb. Secondly, FS fails to derive any universal statements of the iteration of the truth. For example, we can construct in arithmetic a function τ^n with the argument n such that

$$\tau^n(\bar{\sigma}^\top) := \overbrace{T(T \dots T}^{n \text{ times}} (\bar{\sigma}^\top)^\top \dots)^\top,$$

but FS cannot derive $\forall n T(\tau^n(\bar{0} = \bar{0}^\top))$. In other words, the iterability of type-free truth is only realized in a limited way in FS.

27.4.3 Non-classical truth with truth-gaps

Perhaps the most popular type-free axiomatic theory of truth is Feferman’s system KF, which is short for *Kripke-Feferman* and was proposed under a different designation in Feferman (1991) and unpublished work dating back to the 1970s. In this subsection, we will present KF and related theories.

¹⁵ Friedman and Sheard (1987) were the first to consider this theory, though under a different axiomatization.

The truth of FS is classical in the sense that every sentence is true or false (where the falsity of a sentence is identified with the truth of its negation). In other words, FS allows no truth-gap or glut. As we have seen, for the sake of this bivalence of truth, FS sacrifices the iterability of truth.

There is an alternative route toward a consistent theory of type-free truth. Namely, we keep the iterability of truth but allow truth gaps: we allow sentences to be neither true nor false. This strategy naturally leads us to a revision of the classical bivalent evaluation schema. In typical settings, we need to give a new evaluation rule about how the truth and falsity of compositional sentences are determined; for instance, we must determine when $A \wedge B$ is evaluated to be true or false under the possibility that A or B may be neither true nor false. This does not necessarily mean revising the overall logic in which the axiomatic system is formulated; but here we only consider revising the conception of truth represented by the predicate T from the classical bivalent one to non-classical one; in other words, we only change the inner logic while keeping the outer logic classical.

One prominent alternative to the classical bivalent evaluation schema is the so-called strong Kleene evaluation schema. In addition to the classical semantic status of truth and falsity it admits a further third possibility for the semantic status of a sentence, which is usually interpreted as the absence of a definite truth value (*true or false*). The strong Kleene schema is represented by the following tables:

\neg		\wedge	T	-	F
T	F	T	T	-	F
-	-	-	-	-	F
F	T	F	F	F	F

Here “T” and “F” denote the truth and falsity respectively and “-” represents the absence of a definite truth value (*true or false*). The content of the truth tables can be expressed in the following way:

- (a) $\neg\phi$ is true (or false) iff ϕ is false (true, resp.).
- (b-1) $\phi \wedge \psi$ is true iff both ϕ and ψ are true.
- (b-2) $\phi \wedge \psi$ is false iff either ϕ or ψ is false.

Universal quantification is evaluated as infinitary conjunction:

- (c-1) $\forall x\phi(x)$ is true iff all its substitution instances $\phi(t)$ are true
- (c-2) $\forall x\phi(x)$ is false iff at least one of its substitution instances $\phi(t)$ is false.

Here, the truth and falsity of compound sentences are simultaneously defined and the definitional clauses for them interact and depend on each other. Since in the case of the classical bivalent evaluation being false simply means not being true, we have only to postulate compositional axioms for truth as in CT and FS. However, in the present case

of the strong Kleene evaluation, the falsity of a compositional sentence cannot be defined away in terms of the truth of its components. Due to the clause (a) we can still define the falsity of a sentence by the truth of its negation as before, but we don't equate being false with being not true, and thus the compositional axioms for falsity are not derivable only from those for truth (i.e. (a), (b-1) and (c-1)). Hence, we need to add the compositional axioms for falsity separately from those for truth so as to accommodate the strong Kleene evaluation schema in the system.

So we have the following compositional axioms for the non-classical truth and falsity with the strong Kleene evaluation schema:

- κ1 A sentence of the form $s = t$ is true iff s and t coincide in their values.
- κ2 A sentence of the form $s = t$ is false iff s and t differ in their values.
- κ3 An \mathcal{L}_T -sentence of the form $\neg\phi$ is true (or false) iff ϕ is false (true, resp.).
- κ4 An \mathcal{L}_T -sentence of the form $\phi \wedge \psi$ is true iff ϕ and ψ are true.
- κ5 An \mathcal{L}_T -sentence of the form $\phi \wedge \psi$ is false iff ϕ or ψ is false.
- κ6 An \mathcal{L}_T -sentence $\forall x \phi(x)$ is true iff all substitution instances $\phi(t)$ are true.
- κ7 An \mathcal{L}_T -sentence $\forall x \phi(x)$ is false iff at least one substitution instance $\phi(t)$ is false.

Falsity of a sentence is understood as the truth of its negation. Alternatively, we can add falsity as a new primitive symbol to the language.

Similarly to the compositional axioms, we separately introduce the axioms for iteration of truth and falsity. The axioms are of the following forms:

- κ8 An \mathcal{L}_T -sentence of the form Tt is true (false) iff the value of t is a true (false) sentence.
- κ9 An \mathcal{L}_T -sentence of the form Ft is true (false) iff the value of t is a false (true) sentence.

Now, let $\text{KF} \upharpoonright$ be the theory PA plus κ_1 – κ_9 .¹⁶ The full theory KF, together with $\text{KF} \upharpoonright$ -axioms, contains all induction axioms for arbitrary formulae of \mathcal{L}_T . Since we axiomatize a non-classical notion of truth in the classical system KF, the type-free analogue of the axiom c2, that is, F2, is not postulated in $\text{KF} \upharpoonright$ and KF; indeed, it is inconsistent with the other axioms of $\text{KF} \upharpoonright$. In the presence of the axioms κ_8 and κ_9 for iteration of truth and falsity, KF is invulnerable to the aforementioned criticism against FS on the iterability of truth; KF can express the iterability of truth in various forms; for example, KF derives $\forall n T(\tau^n(\ulcorner 0 = 0 \urcorner))$ and $\forall n [T(\tau^n(\ulcorner \neg 0 = 0 \urcorner)) \leftrightarrow T \ulcorner \neg 0 = 0 \urcorner]$, where τ^n is the function defined in the last section.

By definition, $\text{KF} \upharpoonright$ and KF both include PA. As for its deductive power, KF turns out to be a fairly strong theory. KF proves the consistency of FS (and thus that of CT a fortiori). As a matter of fact, KF can even derive the consistency of certain transfinite iterations of typed-truth. Nonetheless, KF is still safe in the sense that it is predicatively justifiable; that is, it is reducible to some predicative mathematics and thus justifiable by predicative

¹⁶ Feferman (1991) employed a predicate F for falsity separately from the truth predicate T ; but it can be defined in terms of T and negation; cf. Halbach (2011); Fujimoto (2010).

means; for the explication of predicativism, see Feferman (2005). In contrast to KF, $\text{KF} \upharpoonright$ is known to be conservative over PA for arithmetical sentences. This means that $\text{KF} \upharpoonright$ is as safe as PA from the viewpoint of consistency.

The KF-axioms are motivated by the notion of so-called *grounded* truth, and this idea is best represented by the following passage of Feferman (1991):

[T]ruth or falsity is *grounded* in atomic facts from the base language \mathcal{L} , i.e. can be determined from such facts by evaluation according to the rules of truth for the connectives and quantifiers, and where statements of the form $[T'A]$ are evaluated to be true (false) only when A itself has already been verified (falsified). Feferman (1991: 18)

Kripke (1975) first tried to formalize this conception of truth in semantic (or model-theoretic) form. Feferman (1991) later gave an axiomatic version of this conception.

The system KF can be regarded as an axiomatic recast of Kripke's theory: given any (classical) model \mathfrak{M} of KF, the extension of the truth predicate T in \mathfrak{M} is at the same time the extension of the truth predicate in some strong Kleene semantics for \mathcal{L}_T constructed by means of Kripke's theory (i.e. a certain fixed-point of a Kripkean operator); furthermore, in the special case in which the arithmetical part (i.e. \mathcal{L}) is interpreted by the standard model \mathbb{N} , arbitrary Kripkean strong Kleene semantics can be conversely turned into a (classical) model of KF. Thus KF is sometimes called an "axiomatization" of Kripke's theory of truth.¹⁷

There are some important variants and extensions of KF. Cantini (1989) considers adding the axiom of consistency or completeness to KF, which respectively express "no sentence is true and false at the same time" and "every sentence is either true or false." It is known that adding exactly one of them to KF results in a consistent theory, though adding both to KF yields a contradiction. In the current setting in which we allow truth-gaps, the axiom of consistency can be fairly naturally added to KF, while the axiom of completeness rather conforms to another setting where we allow truth-gluts (i.e. sentences true and false at the same time) instead of truth-gaps; for further discussion, we refer the reader to Halbach (2011). Feferman (1991) suggests a strengthening of KF by augmenting it with a special inference rule which corresponds to the Bar Rule in second order arithmetic; the resulting system is called the *schematic reflective closure* (of a given base theory); this approach would bring even stronger theory of self-applicable truth.

The background idea of KF explained in the cited passage of Feferman does not restrict our choice of the evaluation rule for truth to a strong Kleene one. We can adopt other evaluation rules and consider "axiomatizations" of Kripke's theory of truth based on those schemata. In some cases, the way we obtain the theory KF for strong Kleene

¹⁷ To be precise, KF is not an "axiomatization" of the *grounded* truth in Kripke's sense. For Kripke, grounded truth is particularly defined as the minimal fixed-point (among many other fixed-points) of his construction. However, as we have seen, KF takes other fixed-points into consideration as well. Reflecting on this, Burgess (2009) proposes adding an axiom schema which expresses the minimality of the truth predicate as a fixed-point in Kripke's construction. Also there is a more general issue on "axiomatization" of semantic theories of truth: what does it formally and precisely mean for an axiomatic theory of truth to "axiomatize" a given semantic theory of truth? See Fischer et al. (2015) for more discussion.

evaluation is directly applicable to other evaluations. Feferman (1991) himself suggests alternatively using weak Kleene evaluation, for example. The system WKF, formulated in the same spirit as KF but based on weak Kleene evaluation, is studied in Fujimoto (2010) and shown to have the same deductive strength as KF; as is also expected, WKF “axiomatizes” Kripke’s theory of truth with weak Kleene evaluation in the same sense as we have explained for the case of KF and strong Kleene evaluation schema.¹⁸ Of course, there are many alternatives other than weak or strong Kleene evaluation.¹⁹

27.4.4 Theories with the domain of significance

A very natural reaction to the liar paradox is probably to say that the liar sentence is meaningless and has no significance. One might thereby conclude that the paradoxes are simply caused by using the truth predicate in an impermissible way. Particularly in the liar case, inconsistency comes from the invalid application of the truth predicate to a meaningless sentence. This leads to the idea that paradoxes could be resolved by prohibiting such illegitimate uses and applications of the truth predicate. We cannot meaningfully apply the truth predicate to some sentences any more than we cannot meaningfully apply the predicate “has a negative electric charge” to the Dow Jones Industrial Average.

The theory of truth based on this view would have two predicates T and D representing truth and its domain of significance, and then the principles concerning truth, such as T -biconditionals or compositional axioms, are restricted to the elements already established as being in D ; thus, in the case of the truth predicate, D consists of the significant sentences and those principles are restricted to them. Accordingly, we postulate the following axioms for this conception of type-free truth:

- D1 For a significant sentence of the form $s = t$, it is true iff s and t coincide in their values.
- D2 For a significant sentence of the form $T\ulcorner\sigma\urcorner$, it is true iff σ is true.
- D3 For a significant sentence of the form $\neg\phi$, it is true iff ϕ is not true.

¹⁸ To be precise, we need an extra condition on the language \mathcal{L} of arithmetic. In the present contribution, we don’t specify how the language of arithmetic should be formulated because most of the arguments generally apply regardless of its formulation. However, the result cited here, that is, that WKF has the same deductive strength as KF, requires that \mathcal{L} contains a sufficiently rich collection of function symbols; in Fujimoto (2010), \mathcal{L} is assumed to contain function symbols for all primitive recursive functions.

¹⁹ For instance, Kripke (1975) considers adopting van Fraassen’s supervaluational evaluation schema instead. In contrast to a strong or weak Kleene case, however, it is rather difficult to apply Feferman’s way of formulating KF for an “axiomatization” of Kripke’s theory with supervaluation schema, and some different technique is needed; this is partially because supervaluation is much more complicated and non-compositional, in the sense that the truth and falsity of sentences are not necessarily determined by those of their subformulae. Such an “axiomatization” VF was given by Cantini (1990).

- D4 For a significant sentence of the form $\phi \wedge \psi$, it is true iff both ϕ and ψ are true.
- D5 For a significant sentence of the form $\forall x \phi(x)$, it is true iff all its substitution instances $\phi(t)$ are true.

These axioms D1–D5 still tell us nothing about which sentences are significant or meaningfully applicable with respect to the truth predicate T . As a trivial example, we can consistently add “no sentence is true” and “no sentence is significant” to them as extra axioms. Thus, the principal task of this approach is to determine an adequate domain of significance for the truth predicate and to give an axiomatic characterization of that domain.

Feferman (2008) presented a theory of truth based on this idea of truth with its domain of significance.²⁰ Besides D1–D5, Feferman argues the domain D of significance in the case of truth consists of *determinate* sentences, that is, sentences with a definite truth value, true or false. In addition, Feferman further requires that D must be *strongly compositional* in the sense that a compound sentence is in D iff all substitution instances of its direct subformulae by (meaningful) terms belong to D . The resulting axioms are accordingly of the following form:

- D6 All sentences of the base language are significant.
- D7 A sentence ϕ is significant iff $T\ulcorner\phi\urcorner$ is significant.
- D8 A sentence of the form $\neg\phi$ is significant iff ϕ is significant.
- D9 A sentence of the form $\phi \wedge \psi$ is significant iff both ϕ and ψ are significant.
- D10 A sentence of the form $\forall x \phi(x)$ is significant iff all its substitution instances $\phi(t)$ are significant.
- D11 A sentence is significant iff it is either true or false.

Let us define a theory $DT\upharpoonright$ as PA plus D1–D11, and then DT is defined as being the result of augmenting $DT\upharpoonright$ with all induction axioms for \mathcal{L}_T . This is shown by straightforward meta-induction on formulae that the T-biconditional is valid for every significant sentence in $DT\upharpoonright$ and DT. We note that this version of DT is different from Feferman’s original theory in one important respect, but this difference has little bearing on our current purpose of explaining how to formulate a theory of type-free truth based on the conception of truth at issue here; as a matter of fact, their deductive strengths are the same.²¹

²⁰ Independently of Feferman, Reinhardt (1986) interpreted KF as a theory of meaningful applicability motivated by a very similar consideration. Like Feferman, he identified the domain D of significance of truth with the set of determinate sentences. He then investigated how this predicate D for the domain of significance behaves and what conditions it meets within KF.

²¹ In his original formulation of DT, Feferman treats the conditional “ \rightarrow ” separately from its ordinary definition in terms of negation and conjunction and postulates special truth-theoretic axioms for the conditional independently of those for negation and conjunction. He does so because some desirable statements, such as “it is true that all the theorem of PA is true,” are not derivable in our current formulation of DT, whereas they are derivable in his original formulation. However, there is a trade-off for the sake of the derivability of those sentences; namely, strong compositionality only partially holds in Feferman’s original formulation.

DT and KF have distinct background motivations and their axioms appear to bear little resemblance. However, these two types of axiomatization are in fact closely related. For example, DT (in our current formulation) turns out to be identical with the theory WKF plus Cantini's consistency axiom (see 27.4.3); for more detailed discussion, see Fujimoto (2010).

27.5 FURTHER PERSPECTIVES

In the previous sections we have sketched the definitions of some axiomatic theories of truth with Peano arithmetic as base theory and formulated in classical logic.

As pointed out above, axiomatic truth theories can be formulated over other base theories. In particular, axiomatic truth theories can be formulated over very strong base theories that include set theory and any of our standard mathematical theories.²² In order to do so, some adjustments are required or advisable. For instance, it might be preferable to axiomatize a binary predicate for satisfaction rather than a unary one for truth.

If the axiomatization of the base theory contains a schema, special caution has to be exercised. We have already mentioned the induction schema of Peano arithmetic. In set theory the schemata of separation and replacement play an analogous role. However, it is not always possible to extend a schema of the base language. Halbach (2006) showed that applying the truth axioms of $TB\upharpoonright$ to a base theory with an extended axiom schema by necessity leads to an inconsistency. Generally, the problems faced by simultaneous axiomatizations of such notions as necessity, knowability, analyticity, and many others have been very little investigated, but obviously theories in which necessity, truth, other notions, and their interaction can be investigated are of the highest philosophical importance.²³ Many philosophers seem so far to be happy with hierarchical approaches whereby, for instance, necessity is expressible in the object-language (for instance, by the modal operator \Box) while the notion of truth is definable only at a higher level in the metalanguage, so that the necessity of statements containing the truth predicate cannot be expressed. In such a setting, the question whether claims involving truth such as the T-biconditionals cannot be addressed, because they cannot be formulated.

There are many further issues that arise when truth theories are applied to different base theories, and many of them have not been investigated to date.

Another line of research we can only touch upon concerns axiomatizations of truth in non-classical logics. There is no shortage of proposals to solve the semantic paradoxes by abandoning classical logic. Typically, the logic of the base theory is kept classical and the laws of classical logic are dispensed with only for formulae containing the truth

²² See Fujimoto (2012) for a study of theories of truth over set theory.

²³ See Halbach and Welch (2009) for an overview.

predicate. For some time solutions of the paradoxes relying on paraconsistent logics have been propagated.²⁴ More recently, Field (2008) has provoked an increased interest in deductive systems for type-free truth in non-classical settings. However, in many cases no precise formal axiom system has been formulated (as opposed to a semantics) and very little is known about the formal properties of axiomatizations of truth in non-classical logics. In particular, very little is known about the strength of non-classical axiomatizations of truth.²⁵ It seems to us that the value of non-classical approaches can only be assessed once a better understanding of the proof-theoretic consequences of giving up classical logic have been obtained.

REFERENCES

- Belnap, N. and A. Gupta (1993). *The Revision Theory of Truth*. Cambridge, MA: MIT Press.
- Burgess, J. P. (2009). "Friedman and the Axiomatization of Kripke's Theory of Truth," in "Foundational Adventures: Essays in Honor of Harvey M. Friedman," paper delivered at the Ohio State University conference in honor of the sixtieth birthday of Harvey Friedman, <http://foundationaladventures.com/>
- Cantini, A. (1989). "Notes on Formal Theories of Truth," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 35: 97–130.
- Cantini, A. (1990). "A Theory of Formal Truth Arithmetically Equivalent to ID₁," *Journal of Symbolic Logic* 55: 244–59.
- Cieśliński, C. (2007). "Deflationism, Conservativeness and Maximality," *Journal of Philosophical Logic* 36: 695–705.
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Enayat, A. and A. Visser (2015). "New Constructions of Satisfaction Classes," in T. Achourioti, H. Galinon, K. Fujimoto, and J. Martinez-Fernandez (eds.), *Unifying the Philosophy of Truth*. Dordrecht: Springer, 321–35.
- Feferman, S. (1991). "Reflecting on Incompleteness," *Journal of Symbolic Logic* 56: 1–49.
- Feferman, S. (2005). "Predicativity," in S. Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*. Oxford: Oxford University Press, 590–624.
- Feferman, S. (2008). "Axioms for Determinateness and Truth," *Review of Symbolic Logic* 1: 204–17.
- Field, H. (2008). *Saving Truth From Paradox*. Oxford: Oxford University Press.
- Fischer, M., V. Halbach, L. Stern, and J. Speck (2015). "Axiomatizing Semantic Theories of Truth?" *Review of Symbolic Logic* 8: 257–78.
- Friedman, H. and M. Sheard (1987). "An Axiomatic Approach to Self-Referential Truth," *Annals of Pure and Applied Logic* 33: 1–21.
- Fujimoto, K. (2010). "Relative Truth Definability," *Bulletin of Symbolic Logic* 16: 305–44.
- Fujimoto, K. (2012). "Classes and Truths in Set Theory," *Annals of Pure and Applied Logic* 163: 1484–1523.

²⁴ See Beall and Ripley (ch. 28 in this volume).

²⁵ Halbach and Horsten (2006) provide a proof-theoretic analysis of an axiomatization of Kripke's theory in strong Kleene logic and show that it is much weaker than its counterpart KF in classical logic.

- Halbach, V. (1994). "A System of Complete and Consistent Truth," *Notre Dame Journal of Formal Logic* 35: 311–27.
- Halbach, V. (1999). "Conservative Theories of Classical Truth," *Studia Logica* 62: 353–70.
- Halbach, V. (2006). "How Not to State the T-sentences," *Analysis* 66: 276–80. Correction of printing error in vol. 67: 268.
- Halbach, V. (2009). "Reducing Compositional to Disquotational Truth," *Review of Symbolic Logic* 2: 786–98.
- Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press.
- Halbach, V. and L. Horsten (2006). "Axiomatizing Kripke's Theory of Truth," *Journal of Symbolic Logic* 71: 677–712.
- Halbach, V. and P. Welch (2009). "Necessities and Necessary Truths: A Prolegomenon to the Metaphysics of Modality," *Mind* 118: 71–100.
- Heck Jr., R. G. (1997). "Tarski, Truth, and Semantics," *The Philosophical Review* 106: 533–54.
- Horwich, P. (1990). *Truth*, 1st edn. Oxford: Basil Blackwell.
- Kotlarski, H., S. Krajewski, and A. Lachlan (1981). "Construction of Satisfaction Classes for Nonstandard Models," *Canadian Mathematical Bulletin* 24: 283–93.
- Kripke, S. (1975). "Outline of a Theory of Truth," *Journal of Philosophy* 72: 690–712. Reprinted in Martin (1984).
- Lachlan, A. (1981). "Full Satisfaction Classes and Recursive Saturation," *Canadian Mathematical Bulletin* 24: 295–7.
- Leigh, G. E. (2013). "Conservativity for Theories of compositional Truth Via Cut Elimination." 1308.0168 [math.LO]; <http://arxiv.org/abs/1308.0168>
- Leitgeb, H. (2007). "What Theories of Truth Should Be Like (But Cannot Be)," *Blackwell Philosophy Compass* 2: 276–90.
- Martin, R. L. (ed.) (1984). *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press.
- McGee, V. (1985). "How Truthlike Can a Predicate Be? A Negative Result," *Journal of Philosophical Logic* 14: 399–410.
- McGee, V. (1992). "Maximal Consistent Sets of Instances of Tarski's Schema (T)," *Journal of Philosophical Logic* 21: 235–41.
- Mostowski, A. (1950). "Some Impredicative Definitions in the Axiomatic Set-Theory," *Fundamenta Mathematicae* 37: 111–24.
- Reinhardt, W. (1986). "Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth," *Journal of Philosophical Logic* 15: 219–51.
- Tarski, A. (1935). "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica Commentarii Societatis Philosophicae Polonorum* 1: 261–405; trans. J. H. Woodger as "The Concept of Truth in Formalized Languages" in Tarski (1956: 152–278); page references are given for the translation.
- Tarski, A. (1956). *Logic, Semantics, Metamathematics*, trans. J. H. Woodger. Oxford: Clarendon Press.
- Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

CHAPTER 28

NON-CLASSICAL THEORIES OF TRUTH

JC BEALL AND DAVID RIPLEY

THIS chapter attempts to give a brief overview of *non-classical*(-logic) theories of truth. Due to space limitations, we follow a victory-through-sacrifice policy: sacrifice details in exchange for clarity of big-picture ideas. This policy results in our giving all too brief treatment to certain topics that have dominated discussion in the non-classical-logic area of truth studies. (This is particularly so of the “suitable conditional” issue: section 28.4.3.) Still, we present enough representative ideas that one may fruitfully turn from this chapter to the more detailed cited works for further study. Throughout—again, due to space—we focus only on the most central motivation for standard non-classical-logic-based truth theories: namely, truth-theoretic paradox (specifically, due to space, the liar paradox).

Our discussion is structured as follows. We first set some terminology concerning theories and logics; this terminology allows us to frame the discussion in a broad-but-clean fashion. (On the logic side, we present a very basic sequent system for truth and negation—and nothing more.) We then present a stripped-down version of the liar paradox. The paradox, as we set it up, turns on four basic rules (not including the truth rules; it’s the job of our target non-classical truth theories to preserve these in unrestricted form): two rules governing negation’s behavior, and two rules governing the “structure” of the validity relation itself. These four rules serve as choice points for the four basic theoretical directions that we sketch. While details, as warned above, are sacrificed for space and big-picture clarity, we hope that the discussion nonetheless charts the main directions of non-classical response to basic truth-theoretic paradox.

28.1 THEORIES AND LOGICS

Since we'll be considering a variety of logics in this chapter, it will help to first have some general tools to work with. We'll adapt, and slightly broaden, the framework of Restall (2013) to this end. For purposes of framing our discussion, we take a *theory* to be a record of both what the given theorist—one who endorses the given theory—accepts and what she rejects (with respect to the given phenomena). Hence, we shall take a theory \mathcal{T} to be a pair $\langle \mathcal{A}, \mathcal{R} \rangle$, where \mathcal{A} and \mathcal{R} contain what an endorser of \mathcal{T} accepts and rejects, respectively.

For some kinds of theory, we might be able to figure out what must be in \mathcal{R} by looking at \mathcal{A} (e.g. each negation in \mathcal{A} might correspond 1–1 to an entry in \mathcal{R}), or vice versa. This is the usual situation with classical theories and classical logic: a classical theorist rejects something iff she accepts its negation. We shall look at two theories that have this feature (see sections 28.5.1 and 28.5.2). On the other hand, some theories may lack this feature: it might be that neither \mathcal{A} nor \mathcal{R} provides sufficient information to derive the other (e.g. negation might fail to track rejection). We shall look at two theories that have this feature (see sections 28.4.1 and 28.4.2).

Each sort of theory we discuss comes with a particular logical approach. We take logics, in “multiple-conclusions” guise, to constrain theories as follows, again herein agreeing with Restall (2005). The argument from premises Γ to conclusions Δ is valid (we write $\Gamma \vdash \Delta$) iff it's out of logical bounds to adopt a theory $\langle \mathcal{A}, \mathcal{R} \rangle$ such that $\Gamma \subseteq \mathcal{A}$ and $\Delta \subseteq \mathcal{R}$. In short: a valid argument rules out certain theories, notably, those theories that accept all of the (valid) argument's premises and reject all of its conclusions.¹

Finally, the logics that we discuss all exhibit two familiar features:

- reflexivity: $A \vdash A$, for any claim A .
- monotonicity: let $\Gamma \subseteq \Gamma'$ and $\Delta \subseteq \Delta'$. If $\Gamma \vdash \Delta$, then $\Gamma' \vdash \Delta'$.

In terms of the interplay with theories, reflexivity tells us that no (logically acceptable) theory $\langle \mathcal{A}, \mathcal{R} \rangle$ involves overlap: $\mathcal{A} \cap \mathcal{R} = \emptyset$. In other words, logic, being reflexive, forbids theorists from both accepting and rejecting one and the same thing.² For monotonicity, define a \mathcal{T} -expanded theory to be any theory $\mathcal{T}' = \langle \mathcal{A}', \mathcal{R}' \rangle$ achieved via superset: $\mathcal{A} \subseteq \mathcal{A}'$ and $\mathcal{R} \subseteq \mathcal{R}'$. Then monotonicity tells us that if (the given) logic rules out a theory \mathcal{T} , it rules out every \mathcal{T} -expanded theory too. In other words, if logic

¹ That one of logic's foundational roles in rational inquiry—particularly rational change in view (as Harman famously puts it) or especially theory expansion—is to proscribe certain theories (or constrain the space of “acceptable” ones) is not only a common idea, but also a very traditional one. Everything we say is compatible with the traditional proscriptive role of logic. We leave open whether logic has any interesting prescriptive role.

² For an approach to paradox that does without this constraint, see Ripley (2013).

rules out accepting Γ while rejecting Δ , then adding more acceptances or rejections won't help.³

28.2 REASONING WITH TRUTH

Throughout the chapter, we use T as our truth predicate, and take $\langle A \rangle$ to be a singular term referring to the sentence A . We simply assume that each sentence A has some such name $\langle A \rangle$, without fussing about *how* $\langle A \rangle$ comes to refer to A ; it can be a quote name, a proper name, a definite description, a Gödel code, or whatever.

There are various familiar principles or rules relating A to $T\langle A \rangle$; we consider three candidates: transparency, the T -schema, and capture and release.

28.2.1 Transparency

Transparency is the principle that A and $T\langle A \rangle$ are intersubstitutable with each other in all non-opaque contexts. Ignoring opaque contexts, then, transparency amounts to everywhere-intersubstitutability. This requires not only that A be equivalent to $T\langle A \rangle$, but also that $A \wedge (\neg B \supset T\langle C \rangle)$ be equivalent to $T\langle T\langle A \rangle \rangle \wedge T\langle \neg T\langle B \rangle \supset C \rangle$, and so on. In short, T s can be added and subtracted willy-nilly, to whole formulas or subformulas. Let formulas that can be obtained from each other by adding and subtracting T s be called *T-variants*.

The notion of equivalence in play can be specified in a few ways. As a constraint on theories, the most natural understanding is this: a theory $\langle \mathcal{A}, \mathcal{R} \rangle$ obeys transparency iff for all A , if $A \in \mathcal{A}$ then every T -variant of A is in \mathcal{A} as well; and if $A \in \mathcal{R}$ then every T -variant of A is in \mathcal{R} as well. This results in A and all its T -variants being equivalent in argument: swapping formulas for their T -variants never makes a valid argument invalid or vice versa.

28.2.2 The T -schema

The *T -schema* is the schema $A \equiv_x T\langle A \rangle$, where \equiv_x is some biconditional or other—typically, in the first instance, a material biconditional (built from negation, disjunction, and conjunction in the usual way). Tarski (1944) offers this schema—in material-biconditional form—as a necessary condition on theories of truth: an adequate theory, he supposes, must have every instance of the T -schema as a theorem.

³ For convenience, we speak of *accepting (set) Γ* and *rejecting (set) Δ* , whereby—note well—we mean *accepting everything in Γ* and *rejecting everything in Δ* , respectively.

On our theory-directed interpretation of theoremhood, there are two ways to understand this: that a theory must accept all instances of the T -schema, or that it must not reject any instances of the T -schema. For our purposes, we don't spend too much time looking at the T -schema, as doing so requires thinking reasonably hard about the status of biconditionals, which we are mostly avoiding here. (See section 28.4.3 for as close as we come to this.)

28.2.3 Capture and release

Capture and *release* are argument forms or “rules of inference” or “extra-logical entailments” (entailments secured by a theory, rather than by logical vocabulary alone). Capture is the rule going from A to $T\langle A \rangle$, the idea being that the truth predicate “captures” the “content” of A , and release is the converse, the rule from $T\langle A \rangle$ to A . On our interpretation, capture rules out any theory that accepts A but rejects $T\langle A \rangle$, and release rules out any theory that accepts $T\langle A \rangle$ but rejects A . Given that logic is reflexive (see above), capture and release follow from transparency. (And if logic enjoys a “deduction theorem,” the T -schema follows from capture and release; however, some of the logics discussed below do not enjoy a deduction theorem. See sections 28.4.1–28.4.3.)

Clearly, transparency, the T -schema, and capture and release have something in common, but they spell it out in different ways. The relations between them are sometimes non-obvious, and always depend on particular features of the background logic. But the core of all three ideas is that A and $T\langle A \rangle$ can stand in for each other in various essential ways. In the non-classical theories sketched below, this core idea remains fixed: at the very least, truth obeys capture and release (if not also being transparent).

28.3 PARADOX AND CLASSICAL LOGIC

In many languages (all natural languages and some formal ones), a sentence can contain a singular term referring to that very sentence itself. For example, the sentence “This very sentence has twenty-three words” includes the singular term “this very sentence”; given a certain context, this term can refer to the sentence itself, rendering it false.

Our main concern in this section is a liar sentence λ which, one way or another, just is $\neg T\langle \lambda \rangle$. In other words, λ is a sentence that says of itself (only) that it is not true. We can produce such a thing in any number of ways; we don't particularly worry about how the trick is pulled here.⁴ The liar causes its trouble by, in some sense, being able to stand in

⁴ For concreteness, we can take λ to be the sentence “The quoted sentence in footnote 4 is not true.”

for its own negation. (The precise sense of *standing in* depends on which properties are taken to govern the truth predicate. We shall, for space reasons, pass over exact details.)

Reasoning classically, we can see that this causes trouble as follows: we cannot reject both the liar and its negation. But since it can stand in for its own negation, this means that we cannot reject both the liar and itself; in other words, we cannot reject it. On the other hand, we cannot accept both the liar and its negation. Since it can stand in for its own negation, this means we cannot accept both the liar and itself; in other words, we cannot accept it. Trouble seems to be afoot.

The classical principles invoked in the foregoing liar-paradoxical reasoning may be summarized as follows: 1) for any sentence, we cannot reject it together with its negation; 2) for any sentence, we cannot accept it together with its negation; 3) if we cannot reject a sentence together with itself, we cannot reject the sentence; 4) if we cannot accept a sentence together with itself, we cannot accept the sentence; and 5) if we cannot accept a sentence and cannot reject it, trouble is afoot.

28.3.1 The liar in sequent form

We proceed to make the given liar-paradoxical argument precise via a Gentzen-style sequent calculus. For our purposes, we needn't worry about conjunction, disjunction, a conditional, quantifiers, or any of that; the rules governing negation, along with the so-called *structural rules*, suffice to cause trouble. (We thus won't consider approaches, like supervaluational or subvaluational approaches, that hinge on fiddling with the behavior of conjunction and disjunction. See McGee 1991; van Fraassen 1968; 1970.)

Our sequents are things of the form $\Gamma \vdash \Delta$, where Γ and Δ are finite “multisets” of formulas. A *multiset* is just like a set, except things can be members of it multiple times, and it matters how many times something is a member (Meyer and McRobbie 1982a; 1982b). Thus, the multiset $[A, A]$ is different from the multiset $[A]$, even though the set $\{A, A\}$ is the same set as $\{A\}$. Multisets do *not* pay attention to order; thus, the multiset $[A, B]$ is the same multiset as $[B, A]$. In an argument with multiple premises, the premises are (as usual) interpreted *conjunctively*; multiple conclusions are dually interpreted *disjunctively*.

In our simple Gentzen system (our *logic*), we take as axioms all sequents of the form $\Gamma, A \vdash A, \Delta$,⁵ and proceed to add three kinds of rule: *contraction* rules, a *cut* rule, and *negation* rules. The first two kinds are *structural*: they don't involve any particular vocabulary. The last kind is *operational*: it tells us what rules negation obeys.

⁵ We set up our axioms with side premises Γ and side conclusions Δ so that all the logics we consider will be *monotonic*: adding premises or conclusions can never make a valid argument invalid. Monotonicity does not seem to be implicated in any of the paradoxes of truth, so we hold it innocent here. (But note that Grišin 1982 finds trouble for monotonicity in a particular naïve set theory.)

First, the contraction rules:

$$\text{Contraction L: } \frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta} \qquad \text{Contraction R: } \frac{\Gamma \vdash A, A, \Delta}{\Gamma \vdash A, \Delta}$$

These tell us that whenever we have multiple occurrences of a premise or a conclusion in a valid argument, the argument remains valid with just a single occurrence of that premise or conclusion. They preserve classical validity, and indeed play a key role in some sequent calculi for classical logic. In terms of theories, they tell us that accepting or rejecting something twice is no stronger than accepting or rejecting it once.

In addition to the two contraction rules (both structural rules), our liar-paradoxical reasoning also includes the following structural rule:

$$\text{Cut: } \frac{\Gamma \vdash A, \Delta \quad \Gamma', A \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'}$$

Cut encodes a generalized form of the *transitivity* of our consequence relation: if B entails A and A entails C , then the cut rule guarantees that B entails C directly; the formula A can be cut out, and argument may proceed directly from B to C . Cut also preserves classical validity in the usual presentations. Unlike the contraction rules, however, the rule of cut is very frequently *eliminable*: it does not expand the stock of provable sequents. It merely provides shortcuts, allowing smaller derivations of some of the very same sequents. In terms of theories, cut is an extensibility condition: it tells us that if some commitments rule out rejecting A , and other commitments rule out accepting it, then combining all of those commitments is ruled out. A theory doesn't have to actually take a stand on A ; cut requires each theory to at least leave open some stand on A .

Finally, our liar-paradoxical argument depends on *operational* rules, namely, rules governing the operator negation. We use the usual classical negation rules:

$$\neg\text{L: } \frac{\Gamma \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta} \qquad \neg\text{R: } \frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \neg A, \Delta}$$

These rules encode the *flip-flop* behavior of classical negation. From the axiom $A \vdash A$, they allow us to prove pivotal sequents:

- *exclusion*: $A, \neg A \vdash$
- *exhaustion*: $\vdash A, \neg A$

Exclusion, derived via $\neg\text{L}$ and reflexivity, tells us that A and its negation may not be accepted together. The second, derived via $\neg\text{R}$ and reflexivity, tells us that A and its negation may not be rejected together.

The foregoing axioms and rules are part of many usual sequent presentations of classical logic; they are enough to reconstruct the above argument for liar-paradoxical trouble, at least given rules governing truth (and the existence of a liar sentence, which we assume). For present purposes, we shall work with capture and release as our rules governing truth, even though the argument can be equally reconstructed with transparency or (given rules for \equiv_x , for some biconditional or other) the T -schema. To accommodate capture and release, we take as additional axioms every instance of the following two schemas:

- capture: $\Gamma, A \vdash T\langle A \rangle, \Delta$
- release: $\Gamma, T\langle A \rangle \vdash A, \Delta$

With all of this in hand, the liar-paradoxical argument may be run as follows.

$$\begin{array}{c}
 \neg R: \frac{p, T\langle \lambda \rangle \vdash \lambda}{p \vdash \lambda, \neg T\langle \lambda \rangle} \quad \neg L: \frac{\lambda \vdash T\langle \lambda \rangle, q}{\lambda, \neg T\langle \lambda \rangle \vdash q} \\
 \text{Contraction R: } \frac{p \vdash \lambda, \neg T\langle \lambda \rangle}{p \vdash \lambda} \quad \text{Contraction L: } \frac{\lambda \vdash T\langle \lambda \rangle, q}{\lambda \vdash q} \\
 \text{Cut: } \frac{p \vdash \lambda \quad \lambda \vdash q}{p \vdash q}
 \end{array}$$

(For the contraction steps, recall that λ just is $\neg T\langle \lambda \rangle$, so we genuinely are contracting two occurrences of the same sentence.) The resulting sequent $p \vdash q$ is absurd: p and q are arbitrary, so a logic that delivers $p \vdash q$ is one according to which anything (whatsoever) entails anything else (whatsoever). This, for our purposes, is completely unacceptable, and so something has to go.⁶ If we take the classical principles appealed to in this argument to be non-negotiable, then it's clear where the adjustment has to be: capture and release (and transparency and the T -schema, as they're implicated in related versions of this argument) must be given up, and so must any theory that entails them. A theory that maintains capture and release, then, must be backed by a logic that does not accept all of $\neg R$, $\neg L$, contraction, and cut. As usual, relaxing logical principles opens space for new theories, theories that would be ruled out if stronger logical principles were held fast.

Here, we discuss four logical options in turn: 1) getting rid of $\neg R$; 2) getting rid of $\neg L$; 3) getting rid of cut; and 4) getting rid of contraction. These four logical options open up different sorts of space for a theory of truth to occupy. As part of our discussion, we also briefly sketch the sort of theory that can live in each kind of logical environment.

⁶ Some accept the conclusion (Azzouni 2006; Kabay 2010), but we won't rebut their arguments here. Our goal is to sketch some of the motivations for non-classical theories, and one such motivation is to avoid this trivialist conclusion.

28.4 OPERATIONAL APPROACHES

Operational approaches are ones that target a particular operator (or class of operators) as the source of liar-paradoxical trouble. In our sample liar derivation above (see section 28.3.1), the only operator involved is negation. The directions of operational approaches that we shall present are those that target negation as the source of trouble—at least initially. (For the potential of additional trouble arising from Curry’s paradox, see section 28.4.3.)

28.4.1 Getting rid of $\neg R$: paracomplete solutions

Getting rid of $\neg R$ amounts to rejecting exhaustion; logical approaches that take this route are known as *paracomplete*. Such logics allow for *paracomplete theories*, where a theory $\mathcal{T} = \langle \mathcal{A}, \mathcal{R} \rangle$ is paracomplete just if both B and $\neg B$ are in \mathcal{R} for some (but not all) sentence(s) B . With respect to the liar, paracomplete theorists reject λ but also reject $\neg\lambda$.

28.4.1.1 *Excluded middle*

Generally, provided that disjunction \vee exhibits standard behavior, paracomplete theorists reject *excluded middle* in the form

$$B \vdash A \vee \neg A$$

This is not to say that paracomplete theorists reject *all instances* of $A \vee \neg A$. Such theorists might think—for extra-logical, certain theory-specific reasons—that, for some specific fragment of the language (e.g. T -free fragment, physics, some such), all instances of $A \vee \neg A$ hold (Field 2008). But they reject that $A \vee \neg A$ is logically true: that it holds via logic alone.

The failure of excluded middle affects the options for T -biconditionals in such theories. This topic is (briefly) discussed below (see section 28.4.3).

28.4.2 Getting rid of $\neg L$: paraconsistent solutions

Getting rid of $\neg L$ amounts to rejecting exclusion; logical approaches that take this route are known as *paraconsistent*. Such logics allow for *paraconsistent theories*, where a theory $\mathcal{T} = \langle \mathcal{A}, \mathcal{R} \rangle$ is paraconsistent just if both B and $\neg B$ are in \mathcal{A} for some (but not all) sentence(s) B .⁷

⁷ In this chapter we ignore—with reluctance!—the distinction drawn in the literature between “paraconsistency” on the one hand and “dialetheism” or “glut theories” on the other. While this distinction is important much of the time, for present purposes it would simply distract.

28.4.2.1 *Explosion*

Generally, provided that conjunction \wedge exhibits standard behavior, paraconsistent theorists reject *explosion* in the form

$$A \wedge \neg A \vdash B$$

This is not to say that paraconsistent theorists accept *all instances* of $A \wedge \neg A$. Such theorists might think—for extra-logical, certain theory-specific reasons—that, for some specific fragment of the language (e.g. T -free fragment, physics, some such), all instances of $A \wedge \neg A$ fail to hold (Beall 2009). But they reject that $A \wedge \neg A$ is logically untrue: that it fails via logic alone.

The failure of explosion affects the options for T -biconditionals in such theories—a topic to which we now very briefly turn.

28.4.3 Suitable conditionals and Curry’s paradox

Our given paracomplete and paraconsistent theories wind up with a non-classical material conditional, where a material conditional $A \supset B$ is defined as $\neg A \vee B$.

- Paracomplete: $\not\vdash A \supset A$.
- Paraconsistent: $A, A \supset B \not\vdash B$.

Hence, in either case, the resulting material conditional is often thought to be inadequate for purposes of underwriting the T -biconditionals.⁸ In the paracomplete case, the given conditional detaches (i.e. validates *modus ponens*) but fails to support all instances of the given (material) T -schema: $T\langle A \rangle \supset A$ and its converse can fail. In the paraconsistent case, all instances of the given (material) T -schema hold; however, the given conditional fails to detach.

⁸ One easy way to establish such “inadequacies” is via a common (sound and complete) “semantics” for common such logics—e.g. strong Kleene or K_3 (Kleene 1952; Beall and van Fraassen 2003) and LP (Asenjo 1966; Priest 1979; Beall and van Fraassen 2003). In short: let \mathbb{V} contain all (total) maps v from sentences into $\{1, .5, 0\}$ such that $v(\neg A) = 1 - v(A)$, $v(A \wedge B) = \min\{v(A), v(B)\}$, and $v(A \vee B) = \max\{v(A), v(B)\}$. In the paracomplete K_3 case, we say that $v \in \mathbb{V}$ *satisfies* A just if $v(A) = 1$, and *dissatisfies* A otherwise. In the LP case, we say that $v \in \mathbb{V}$ *satisfies* A just if $v(A) \in \{1, .5\}$, and *dissatisfies* A otherwise. In both cases, we say that $v \in \mathbb{V}$ satisfies a set Γ iff v satisfies each member of Γ , and $v \in \mathbb{V}$ *dissatisfies* Γ iff v *dissatisfies all elements of* Γ . Finally, we may define, for each of the given logics L , “semantic consequence” $_L$ in the foregoing terms: $\Gamma \vdash_L \Delta$ iff there’s no $v \in \mathbb{V}$ that satisfies Γ but *dissatisfies* Δ . Where L is taken to be K_3 , with (dis-)satisfaction defined as above, \vdash_L is paracomplete (as an easy exercise shows); and, dually, \vdash_L is paraconsistent where L is taken to be LP, with (dis-)satisfaction defined as above. (NB: we have actually given what we have elsewhere called K_3^+ and LP^+ , respectively, in order to maintain uniformity with our multiple-conclusion-based discussion in sequent-calculus terms. See Beall (2011; 2013). Strictly speaking, K_3 and LP are the single(ton)-conclusion limits of K_3^+ and LP^+ , so understood.)

As a result of these apparent deficiencies, much of the work in paraconsistent and paracomplete responses to paradox has focused on supplementing such theories with a suitable conditional, one that both detaches and validates all *T*-biconditionals (Beall 2009; Field 2008; Priest 2006; Brady 2006). But the task is difficult. What makes the task particularly difficult is Curry's paradox (Meyer et al. 1979), which involves (conditional) sentences that say of themselves (only) that *if* they are true *then* absurdity is true (e.g. that everything is true).⁹ In the material-conditional case, Curry's paradox is nothing more than a disjunctive version of the liar (e.g. "Either I'm not true or absurdity is true"), which is already treated by standard paraconsistent or paracomplete approaches to the liar. But when a new "suitable conditional" has been added to the mix, Curry's paradox is a distinct—and very, very difficult—problem (Myhill 1975). In fact, Curry's paradox has often been regarded as the hardest obstacle in the path of "para-" solutions to paradox (Beall et al. 2006; Field 2008; Priest 2006).¹⁰

For space reasons, we need omit discussion of the various avenues toward adding detachable, but Curry-paradoxical-safe, *T*-biconditionals to paracomplete and paraconsistent theories (Beall 2009; Brady 2006; Field 2008; Priest 2006). But we should mention a relatively unexplored alternative: simply accept the deficiencies of the material *T*-biconditionals, but respond to them in some other fashion. One approach is to devise a suitable non-monotonic logic, and try to "capture back" as much of the otherwise lost features of the *T*-biconditionals (Goodship 1996; Priest 1991). Another route is to move to a multiple-conclusion logic and an appropriate philosophy thereof (e.g. one that sees the work of "detachment" not in a detachable conditional but instead in extralogical principles that ground the inference from certain premises to certain conclusions) (Beall 2013; 2015). The viability of such approaches remains open.

28.5 SUBSTRUCTURAL APPROACHES

The above approaches work at the level of *operational* rules, in particular the rules governing negation. But classical negation is useful for many purposes. For example, as we've seen above, paracompletists and paraconsistentists alike must reject the usual

⁹ Worth noting here is that in popular paracomplete logics such as strong Kleene, the material conditional fails to enjoy a deduction theorem. Example: $A \vdash A$ but $\not\vdash A \supset A$. On the other (dual) side, with the corresponding (dual) paraconsistent logic LP, the other direction of the deduction theorem fails: $\vdash (A \wedge (A \supset B)) \supset B$ but $A \wedge (A \supset B) \not\vdash B$. In general, for Curry-paradoxical reasons, theories cannot have a deduction theorem for a detachable conditional—at least if the underlying structural rules contain both transitivity and contraction. (See 28.5 for more discussion.)

¹⁰ While we cannot discuss it, we should mention too that Curry's paradox equally confronts "property theories" that purport to accommodate properties corresponding to each meaningful predicate—in short, each meaningful predicate picks out a property exemplified by all and only the objects of which the predicate is true. Having this sort of theory confronts Curry's paradox in the (Russell-like) form of the property exemplified by all and only those things such that if they exemplify themselves, then absurdity follows.

understanding of the relations between acceptance, rejection, and negation: paracomplete theorists *reject* some A without thereby *accepting* $\neg A$, while paraconsistent theorists *accept* some $\neg A$ without thereby rejecting A , and so on. In addition, the paracompleteist loses the law of excluded middle, and the paraconsistentist loses explosion, both familiar and useful principles of inference. Finally, the loss of excluded middle or explosion removes much of the conditional flavor of the classical material conditional. For these reasons, an approach that allows us to proceed without losing so much might be thought superior over the para- accounts.

Here, we briefly outline two *substructural* approaches. These work at the level of the structural rules, so they allow for the maintenance of both $\neg L$ and $\neg R$, restoring much of the usefulness of classical negation and the classical material conditional. But they too are not without costs, as we note below.

28.5.1 Getting rid of cut: nontransitive solutions

The first substructural approach we consider retains the rules of contraction and dispenses with the rule of cut; this results in a *nontransitive* logic. On an approach like this, both of the sequents $p \vdash \lambda$ and $\lambda \vdash q$ are derivable, but without the rule of cut there is no way to derive $p \vdash q$, so the disaster is averted at the very last step.

Nontransitive logics have been advanced in Weir (2005) and Ripley (2013) for handling truth-theoretic paradoxes. They block the problematic derivation, and they do so in a way that allows them to preserve classical operational rules.¹¹ This allows the resulting logical systems to behave quite naturally in a number of ways.

By preserving the classical flip-flop behavior of negation, the nontransitive theorist also preserves the conditional flavor of the material conditional. Nontransitive logics, like the logic ST discussed in Ripley (2013) and Cobreros et al. (2014), can maintain the trinity which the para- approaches, in one way or another, abandon:

- \supset -identity: $\vdash A \supset A$
- \supset *modus ponens*: $A, A \supset B \vdash B$
- deduction theorem: $\Gamma, A \vdash B, \Delta$ iff $\Gamma \vdash A \supset B, \Delta$.

Approaches that focus exclusively on operational rules not only must fail some of these for the material conditional, but in fact must fail some of these for *any* conditional, due to Curry paradox. (Proof: exercise, but use the above rules and standard structural rules, plus release and capture.) This means that nontransitive logics can make do with material conditionals and, in fact, material T -biconditionals: there is no need either to add

¹¹ The system presented in Weir (2005) preserves many, but not all, classical operational rules; the system presented in Ripley (2013) preserves them all. As a result, we focus in this section on the latter system. Note that Weir's approach rejects the validity of $p \vdash \lambda$ and $\lambda \vdash q$; it works in a related but distinct way.

a separate “suitable conditional” or to learn to live with oddly behaved conditionals—unlike in the paracomplete and paraconsistent theories, which, as mentioned in section 28.4.3, must take one of these routes.

There is a reason why nontransitive logics can behave so classically. Recall that cut, unlike contraction, is eliminable in many presentations of (truth-free) classical logic; this means that it plays no essential role in any derivation. Anything that can be derived with it can also be derived without it. As our above liar-based argument shows, this is no longer true when the behavior of truth is accounted for; with capture and release on board, cut makes a genuine difference. However, it only makes a difference to derivations in which capture and release are involved; as a result, one can preserve *every* classically-valid argument in a nontransitive logic. As is shown in Ripley (2013), one can even ensure that all of these arguments extend to cover the full, truth-involving, language.

There is thus a clear sense in which such a nontransitive system is not non-classical: it validates every classically valid argument, in the full vocabulary of the language, including when a truth predicate is present. Nonetheless, the loss of transitivity is at least unfamiliar, and the motivations for adopting such a logic are very similar to many non-classicists’ motivations; there is an equally clear sense in which such an approach *is* non-classical. We won’t bother with the terminological question here.

As we sketched above, the rule of cut amounts to the following constraint on theories: every theory must leave open either accepting A or rejecting it. Ripley takes λ to provide a counterexample to this principle and thus to transitivity. Deriving $\vdash \lambda$ thus tells us that it’s incoherent to reject λ , and deriving $\lambda \vdash$ that it’s incoherent to accept it. The nontransitivist of this stripe must neither accept nor reject λ . This is the theory offered of λ ’s paradoxicality: it cannot be accepted or rejected without incoherence. Unlike the operational approaches, this nontransitive theory maintains the equivalence between accepting $\neg A$ and rejecting A , and between rejecting $\neg A$ and accepting A . Thus, $\neg \lambda$ too must be neither accepted nor rejected. In acceptance, then, this approach is like a paracomplete approach: it accepts neither λ nor $\neg \lambda$. In rejection, it is like a paraconsistent approach: it rejects neither λ nor $\neg \lambda$. However, given our above definitions, this theory is neither paracomplete nor paraconsistent.¹²

28.5.2 Getting rid of contraction: noncontractive solutions

The other sort of substructural approach we’ll consider retains the rule of cut, and does without the rules of contraction. Such an approach is recommended and outlined in Beall and Murzi (2013); Shapiro (2010); and Zardini (2011). On a noncontractive

¹² For a variant nontransitive theory that is both paracomplete and paraconsistent on the present definitions, see Ripley (2013).

approach, one can allow that the sequents $p \vdash \lambda, \lambda$ and $\lambda, \lambda \vdash q$ are derivable, but insist that the sequents $p \vdash \lambda$ and $\lambda \vdash q$ are not; this blocks the derivation of $p \vdash q$.¹³

Moreover, it blocks the derivation in a way that allows for the negation rules and the cut rule to be preserved. This allows the resulting logical systems to behave quite intuitively in a number of ways. By preserving the classical flip-flop behavior of negation, the noncontractive theorist, like the nontransitive theorist, preserves the conditional flavor of the material conditional. Noncontractive logics can thus also maintain all of \supset -identity, *modus ponens*, and the deduction theorem.¹⁴

If $\&$ is the conjunction that reflects the operation of premise combination (multiplicative conjunction; see fn. 14), then it is no longer idempotent on a noncontractive logic; $A \& A$ is stronger than A alone. Similarly, if \vee is the disjunction that reflects the operation of conclusion combination (multiplicative disjunction; see fn. 14), then it too is no longer idempotent; $A \vee A$ is weaker than A alone. It is these differences that are exploited in the noncontractive approach to paradoxes. By arguments similar to those in section 28.3.1, we have both $\vdash \lambda, \lambda$ and $\lambda, \lambda \vdash$ without any uses of contraction. If $\&$ and \vee are as above, this means we have $\vdash \lambda \vee \lambda$ and $\lambda \& \lambda \vdash$; i.e. $\lambda \vee \lambda$ is a logical truth, and $\lambda \& \lambda$ is explosive. Classically, this would be a problem, since classically $A \vee A$ is equivalent to $A \& A$. But noncontractively this is not so; since $\lambda \vee \lambda$ is weaker than $\lambda \& \lambda$, this is no trouble at all.

The noncontractive approach requires us to add subtlety to our account of theories. Recall that for the other approaches we consider, a theory is a pair of *sets*: \mathcal{A} , the things accepted by the theory, and \mathcal{R} , the things rejected by the theory. We then said that $\Gamma \vdash \Delta$ iff it's ruled out to accept everything in Γ and reject everything in Δ . In a noncontractive logic, however, we can have $\Gamma \vdash A, A, \Delta$ without $\Gamma \vdash A, \Delta$: it can be that rejecting A twice is ruled out but rejecting A once is not. This means that, to specify a theory in a noncontractive logic in the corresponding way, we need to keep track of more than *whether* something is accepted or rejected; we also need to keep track of *how many times* it is accepted or rejected.

We do this as follows: a theory is still a pair $\langle \mathcal{A}, \mathcal{R} \rangle$. Now, however, \mathcal{A} and \mathcal{R} are no longer sets; they are rather ω -long *sequences* of sets. We index them with natural numbers for easy reference: thus, $\mathcal{A} = \langle \mathcal{A}_1, \mathcal{A}_2, \dots \rangle$, and $\mathcal{R} = \langle \mathcal{R}_1, \mathcal{R}_2, \dots \rangle$. For any n , \mathcal{A}_n is the set of formulas that the theory in question accepts at least n times, and \mathcal{R}_n is the set of formulas that the theory in question rejects at least n times. Given this set-up, we have $\mathcal{A}_1 \supseteq \mathcal{A}_2 \supseteq \dots$ and $\mathcal{R}_1 \supseteq \mathcal{R}_2 \supseteq \dots$. Now, we can extend our reading of logical consequence

¹³ If we try to use the rule of cut to combine $p \vdash \lambda, \lambda$ and $\lambda, \lambda \vdash q$, we can only cut out a single occurrence of λ from each sequent; we end up with $p, \lambda \vdash \lambda, q$. This is no problem; in fact, it's an axiom!

¹⁴ Whether \supset -contraction is preserved depends on the precise rules used to govern \supset . In the absence of contraction, conjunction, disjunction, and the conditional come in two distinct flavors each; these are sometimes called "additive" and "multiplicative" flavors. (In the presence of both monotonicity and contraction, these two flavors are equivalent.) Noncontractive approaches retain \supset -contraction for the additive \supset , but not the multiplicative.

to noncontractive approaches. We say that $\Gamma \vdash \Delta$ iff no theory can accept each thing in Γ as many times as it appears in Γ and reject each thing in Δ as many times as it appears in Δ .¹⁵

Since the noncontractive approach maintains that $p \vdash \lambda, \lambda$, we have it that no theory can accept p even once and reject λ twice; however, since $p \nvdash \lambda$,¹⁶ it's ok for a theory to accept p once and reject λ once. Similarly, since $\lambda, \lambda \vdash q$, no theory can accept λ twice and reject q , but since $\lambda \nvdash q$, it's ok for a theory to accept λ once and reject q . Thus, the noncontractivist, on this reading, maintains that it's ok for a theory to accept λ and ok for a theory to reject it, so long as it only does one of the two, and only does it once. The natural question at this point is: how can it be that accepting or rejecting something once can be ok when accepting or rejecting it twice is out of bounds?

Actual noncontractivists have not tended to frame their views in terms of bounds on acceptance and rejection, so they have not offered an answer to this precise question. Thus we pause to very briefly sketch a few understandings of noncontractive consequence that have been offered. Zardini (2011) suggests that the liar sentence exhibits a kind of *instability* reminiscent in some ways of a revision theory. The idea is that from a single occurrence of λ one may derive (via the truth rules) $\neg\lambda$, but in the process of doing this the original occurrence of λ was destroyed; thus, we don't have λ and $\neg\lambda$ together, which is a good thing, since $A, \neg A \vdash$. On the other hand, if we have two occurrences of λ , we can use one to derive $\neg\lambda$. This may destroy it, but we still have another copy; we then have both λ and $\neg\lambda$ together, which entails anything, even though the liar on its own does not. Beall and Murzi (2013) suggest thinking of premises as *resources* to be drawn on in the course of a proof. If drawing on a premise uses it up, then again we can see why two occurrences can get us farther than one. In a similar vein, Mares and Paoli (2014) offer a picture of consequence as *information extraction*; if we need to use the information contained in a premise twice, that requires us to actually have the premise twice, on their view.

28.6 CONCLUSION

Classical logic (including cut) seems to rule out the possibility of giving a theory of truth that validates capture and release, or transparency, or the *T*-schema. In this chapter, we've looked at four ways to modify this logical background to open up space for such

¹⁵ Since \vdash is still reflexive and monotonic, we have it that no \mathcal{A}_i can overlap any \mathcal{R}_j ; accepting something any number of times rules out rejecting it any number of times, and vice versa.

¹⁶ The noncontractive theorist had better not accept $p \vdash \lambda$, since then two cuts with the derivable sequent $\lambda, \lambda \vdash q$ would yield the unacceptable $p \vdash q$. Similarly, they had better not accept $\lambda \vdash q$, since then two cuts with the derivable sequent $p \vdash \lambda, \lambda$ would again yield the unacceptable $p \vdash q$. This is why the noncontractive approach quite crucially must go without *both* contraction on the left and contraction on the right; this contrasts with the operational approaches above, which only need to go without a single negation rule each, and can keep the other.

a theory of truth, and looked at the kinds of theory that fit most naturally with each modification. Two of the modifications were to the classical theory of negation; these paracomplete and paraconsistent approaches removed the requirements of exhaustiveness and exclusiveness, respectively. Relaxing exhaustiveness allows for rejecting both the liar and its negation; relaxing exclusiveness allows for accepting them both. Changing the theory of negation has effects on the theory of the material conditional as well, and these effects are a central focus of paracomplete and paraconsistent approaches (see section 28.4.3).

The other two modifications were to structural rules; noncontractive and nontransitive approaches can keep the full classical theory of negation, but must make adjustments elsewhere, either by supposing that two occurrences of the same premise or conclusion amount to more than a single occurrence, or else by supposing that logical consequence is nontransitive. Either way, these substructural solutions owe a theory of logical consequence that can make sense of these adjustments; we've tried to sketch what such theories might look like.

Our discussion has skipped over philosophical arguments for maintaining (unrestricted) capture and release, and also skipped over topics (and common terminology) of "gaps," "gluts," and more. These topics are all important, and have not been dealt with here purely on account of space constraints. We leave such topics to other discussion (Beall and Glanzberg 2008), including much of the work we've cited throughout.

REFERENCES

- Asenjo, F. G. (1966). "A Calculus of Antinomies," *Notre Dame Journal of Formal Logic* 16: 103–5.
- Asenjo, F. G. and J. Tamburino (1975). "Logic of Antinomies," *Notre Dame Journal of Formal Logic* 16: 17–44.
- Azzouni, J. (2006). *Tracking Reason: Proof, Consequence, and Truth*. New York: Oxford University Press.
- Beall, J. (2009). *Spandrels of Truth*. Oxford: Oxford University Press.
- Beall, J. (2011). "Multiple-conclusion LP and Default Classicality," *Review of Symbolic Logic* 4: 326–36.
- Beall, J. (2013). "LP⁺, K3⁺, FDE⁺ and their Classical Collapse," *Review of Symbolic Logic* 6: 742–54.
- Beall, J. (2015). "Free of Detachment: Logic, Rationality, and Gluts," *Noûs* 49: 410–23.
- Beall, J., R. Brady, A. Hazen, G. Priest, and G. Restall (2006). "Relevant Restricted Quantification," *Journal of Philosophical Logic* 35: 587–98.
- Beall, J. and M. Glanzberg (2008). "Where the Paths Meet: Remarks on Truth and Paradox," in P. A. French and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*. Boston: Blackwell, 169–98.
- Beall, J. and J. Murzi (2013). "Two Flavors of Curry's Paradox," *Journal of Philosophy* 110: 143–65.
- Beall, J. and B. C. van Fraassen (2003). *Possibilities and Paradox: An Introduction to Modal and Many-Valued Logic*. Oxford: Oxford University Press.
- Brady, R. (2006). *Universal Logic*, CSLI Lecture Notes 109. Stanford: CSLI Publications.

- Cobrerros, P., P. Égré, D. Ripley, and R. van Rooij (2014). "Reaching Transparent Truth," *Mind* 122: 841–66.
- Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.
- Fine, K. (1975). "Vagueness, Truth and Logic," *Synthese* 30: 265–300.
- Fraassen, B. C. van (1968). "Presupposition, Implication and Self-reference," *Journal of Philosophy* 65: 136–52.
- Fraassen, B. C. van (1970). "Truth and Paradoxical Consequences," in R. Martin (ed.), *The Paradox of the Liar*. Atascadero: Yale University Press, 13–23.
- Goodship, L. (1996). "On Dialetheism," *Australasian Journal of Philosophy* 74: 153–61.
- Grišin, V. N. (1982). "Predicate and Set-Theoretic Calculi Based on Logic Without Contractions," *Mathematics of the USSR-Izvestija* 18: 41–59. (Russian original 1981).
- Kabay, P. (2010). *On the Plenitude of Truth: A Defense of Trivialism*. Saarbrücken: Lambert.
- Kleene, S. C. (1952). *Introduction to Metamathematics*. Princeton: Van Nostrand.
- Mares, E. and F. Paoli (2014). "Logical Consequence and the Paradoxes," *Journal of Philosophical Logic* 43: 439–69.
- McGee, V. (1991). *Truth, Vagueness, and Paradox*. Indianapolis: Hackett.
- Meyer, R. K. and M. A. McRobbie (1982a). "Multisets and Relevant Implication i," *Australasian Journal of Philosophy* 60: 107–39.
- Meyer, R. K. and M. A. McRobbie (1982b). "Multisets and Relevant Implication ii," *Australasian Journal of Philosophy* 60: 265–81.
- Meyer, R. K., R. Routley, and J. M. Dunn (1979). "Curry's Paradox," *Analysis* 39:124–8.
- Mortensen, C. (1995). *Inconsistent Mathematics*. Dordrecht: Kluwer.
- Myhill, J. (1975). "Levels of Implication," in A. R. Anderson, R. Barcan-Marcus, and R. C., Martin (eds.), *The Logical Enterprise*. New Haven: Yale University Press, 179–85.
- Priest, G. (1979). "The Logic of Paradox," *Journal of Philosophical Logic* 8: 219–41.
- Priest, G. (1991). "Minimally Inconsistent LP," *Studia Logica*, 50: 321–31.
- Priest, G. (2006). *In Contradiction*, 2nd edn. Oxford: Oxford University Press. First printed by Martinus Nijhoff in 1987.
- Priest, G. and R. Routley (1989). "First historical introduction: A preliminary history of paraconsistent and dialethic approaches," in G. Priest, R. Routley, and J. Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Munich: Philosophia Verlag, 3–75.
- Restall, G. (2005). "Multiple Conclusions," in P. Hajek, L. Valdes-Villanueva, and D. Westerståhl (eds.), *Logic, Methodology, and Philosophy of Science: Proceedings of the Twelfth International Congress*. London: Kings' College Publications, 189–205.
- Restall, G. (2013). "Assertion, Denial, and Non-Classical Theories," in K. Tanaka, F. Berto, E. Mares, and F. Paoli (eds.), *Paraconsistency: Logic and Applications*. Dordrecht: Springer, 81–100.
- Ripley, D. (2013). "Paradoxes and Failures of Cut," *Australasian Journal of Philosophy* 91: 139–64.
- Shapiro, L. (2011). "Deflating Logical Consequence," *Philosophical Quarterly* 61: 320–42.
- Tarski, A. (1944). "The Semantic Conception of Truth," *Philosophy and Phenomenological Research* 4: 341–76.
- Weir, A. (2005). "Naïve Truth and Sophisticated Logic," in J. Beall and B. Armour-Garb (eds.), *Deflationism and Paradox*. Oxford: Oxford University Press, 218–49.
- Zardini, E. (2011). "Truth Without Contra(di)ction," *Review of Symbolic Logic* 4: 498–535.

CHAPTER 29

CONTEXTUAL THEORIES OF TRUTH AND PARADOX

KEITH SIMMONS

29.1 THE REVENGE LIAR

CONTEXTUAL theories of truth are motivated by a powerful version of the liar paradox, sometimes called the *revenge liar*. We start with the following sentence:

(L) L is not true.

L is semantically defective—so it follows that L is not true. But then we’ve just inferred the liar sentence L—so L says something true, and we’re landed back in paradox. Or consider the sentence:

(P) P does not express a true proposition.

We can infer that, on pain of contradiction, P does not express a (true or false) proposition, and so does not express a true proposition. And from this it follows that P is true. The general pattern here is that the assessment of the liar sentence as semantically defective entails the conclusion that the liar sentence is true.

This is a particularly powerful version of the liar paradox because it seems unavoidable. Surely any approach to semantical paradox *must* declare liar sentences defective in some way or other—liar sentences fail to have a truth value or fail to express a proposition, or they’re not grounded or not stably true or not determinately true, and so on. But then a liar sentence that ascribes to itself one or other of these semantic defects says something true.

Contextual theories take this revenge paradox to be at the heart of the problem raised by the liar. For many other approaches, the revenge liar is a threat that is waiting in the wings. For example, Kripke’s theory of truth appeals to gaps and the notion of groundedness, and so must address the challenge presented by “This sentence is false

or gappy” or “This sentence is false or ungrounded.” The challenge takes the form of a dilemma: either the notion of gappiness or ungroundedness is inexpressible, or a hierarchy of expressively richer languages must be admitted. The first horn is problematic, since Kripke’s theory is expressed in terms of gaps and grounding; and the second horn is problematic for any theory, like Kripke’s, that seeks to avoid a Tarskian hierarchy. Similar dilemmas confront Gupta’s revision theory of truth (consider e.g. “This sentence is not stably true”) and Field’s paracomplete theory (e.g. “This sentence is not hyper-determinately true”).¹

But for contextual theories, the revenge liar is the appropriate place to begin. In the course of revenge reasoning the liar sentence shifts in semantic status from defective to true. We assess the sentence *L* as defective, and then as true; we reason that *P* fails to express a proposition, and go on to reason that it’s true. It seems that the reasoning in each case is natural and valid, and the shifts in semantic status are genuine phenomena that call for explanation. For the contextualist about truth, these shifts indicate that truth is *context-dependent* in some way—context plays a crucial role in fixing the semantic status of liar sentences.

In section 29.2, I set out a revenge discourse. In section 29.3, I review recent work by philosophers, semanticists, and linguists on the kinematics of context-change. In section 29.4, I show how contextual accounts of the revenge liar can draw on this literature. In sections 29.5–29.8, I review the two main kinds of contextualist theory. According to the first, the truth predicate itself is context-sensitive (see e.g. Burge 1979; Koons 1992; and Simmons 1993). According to the second, the context-dependence of the liar sentence is derived from a more general source, the context-dependence of the background domain (see e.g. Parsons 1974; Glanzberg 2001; Barwise and Etchemendy 1984). In sections 29.5 and 29.6, I turn to theories of the first kind, and in sections 29.7 and 29.8, to theories of the second kind. In section 29.9, I consider two major challenges to all contextual theories of truth.

29.2 A REVENGE DISCOURSE

Contextualists about truth take revenge liar discourses as their principal data. Here is one example of a revenge discourse. Start with:

(R) The sentence written at the third line of Section 2 of this article is not true.

¹ For Kripke’s theory, see Kripke (1975); for Gupta’s revision theory, see Gupta (1982); and for Field’s paracomplete theory, see Field (2008). For critical discussion of Kripke and Gupta, see e.g. Simmons (1993: chs 3 and 4); and for critical discussion of Field, see e.g. Priest (2005: 44–6); Beall (2005: 23–4); Yablo (2003: 328–9); Scharp (2008). Field seems to prefer the second horn of the dilemma—see Field (2008: chs 22 and 23).

Suppose R is true—then it isn't. Contradiction. Suppose R is false—then it's not true. But that's what R says of itself—so R is true. Contradiction, again. So we reach the conclusion that R is semantically defective. But that is not the end of the matter; our reasoning need not stop here. If we look carefully at ways in which the reasoning can be extended, three distinct phenomena emerge, phenomena that call for explanation. To give them labels, these phenomena are *Repetition*, *Rehabilitation*, and *Oscillation*.

Repetition. Given that R is defective, we may infer:

(S) The sentence written at the third line of Section 2 of this article is not true.

S is validly inferred from the truth that R is defective—so S is true. This shows that we can *repeat* the words of a liar sentence and thereby produce a truth.

Rehabilitation. We may continue: since R and S say exactly the same thing, and S is true, we may conclude that R is also true. Here we *rehabilitate* the liar sentence R—we declare R true, because it's true that the sentence written at the third line of Section 2 of this article is not true, and that's exactly what R says.

Oscillation. We've now evaluated R as true. But then, given what R says, it follows that R is not true. But if R is not true, then that's what it says, so it's true. And so on. We now *oscillate* between opposing values for R.

The reasoning that generates *Repetition*, *Rehabilitation*, and *Oscillation* appears natural and intuitive. In each case, we are led to a surprising or troubling conclusion: that two tokens of the same type with the same linguistic meaning differ in semantical status (repetition); that the semantic status of a single sentence shifts from defective to true (rehabilitation); that one and the same sentence has opposing truth values (oscillation). Contextual accounts seek to provide an adequate explanation of these phenomena while respecting the naturalness and intuitiveness of the reasoning—this valid reasoning is not to be blocked in an artificial way.² The three phenomena are to be regarded as significant data which require explanation.

29.3 CONTEXT AND CONTENT

We're familiar with the idea that context acts on content, as with indexical terms like "I" and "now." But increasingly it is recognized that the reverse direction holds as well: *content acts on context*. Stalnaker writes,

[C]ontext constrains content in systematic ways. But also, the fact that a certain sentence is uttered, and a certain proposition expressed, may in turn constrain or alter

² Scharp for instance replaces the concept of truth (which he regards as inconsistent) by the concepts of *ascending truth* and *descending truth*, and as a result cannot allow the intuitive revenge reasoning to go through (see e.g. Scharp 2007).

the context . . . There is thus a two-way interaction between contexts of utterance and contents of utterances.³

At a given point in a discourse, the context will in part depend on what has been said before. For example, the context may change as new information is added. According to Stalnaker, the connection between context and available information is very tight indeed. Stalnaker writes,

I propose to identify a context (at a particular point in a discourse) with the body of information that is presumed, at that point, to be common to the participants in the discourse.⁴

To put it another way, a context is to be represented by the shared presuppositions of the participants—or the “common ground,” to use a phrase from Grice.⁵ As new utterances are produced, and new information is made available, the context changes. For a specific example, consider the speech act of assertion: “Any assertion changes the context by becoming an additional presupposition of subsequent conversation.”⁶

The shared presuppositions of conversants also figure in David Lewis’s account of context-change. Lewis introduces the notion of a *conversational score*.⁷ Following Stalnaker, he identifies the set of shared presuppositions of the participants (at a given stage of a conversation) as one component of the conversational score. “Presuppositions can be created or destroyed in the course of a conversation”⁸—and as the set of presuppositions changes, the conversational score changes. Of course, the notion of conversational score is a vivid way of capturing the notion of context. A change in the set of presuppositions is a change of context.

³ Stalnaker (1975) in Stalnaker (1999: 66). Isard puts it this way: “communications do not merely depend on the context for their interpretation, they change that context” (Isard 1975).

⁴ Stalnaker (1988) in Stalnaker (1999: 98). This is a repeated theme in Stalnaker’s writings; e.g. on p. 6 of his introduction to Stalnaker (1999), he writes: “. . . a context should be represented by a body of information that is presumed to be available to the participants in the speech situation.”

⁵ Grice (1989).

⁶ Appendix to Stalnaker (1975) in Stalnaker (1999: 77). In a similar vein, Stalnaker writes: “. . . the essential effect of an assertion is to change the presuppositions of the participants in the conversation by adding the content of what is asserted to what is presupposed” (Stalnaker 1978 in Stalnaker 1999: 86).

⁷ See Lewis (1979). The analogy is with a baseball score. A baseball score for Lewis is composed of a set of seven numbers that indicate, for a given stage of the game, how many runs each team has, which half of which innings we’re in, and the number of strikes, balls, and outs. Notice that correct play depends on the score—what is correct play after two strikes differs from what is correct play after three strikes. Similarly for conversations: the correctness of utterances—their truth, or their acceptability in some other respect—depends on the *conversational* score. Lewis continues, “Not only aspects of acceptability of an uttered sentence may depend on score. So may other semantic properties that play a role in determining aspects of acceptability. For instance, the constituents of an uttered sentence—subsentences, names, predicates, etc—may depend on the score for their intension or extension” (Lewis 1979: 345).

⁸ Lewis (1979: 339).

Another component of the conversational score, according to Lewis, is the *standard of precision* that is in force at a given stage of the discourse. Suppose I say “France is hexagonal.” If you have just said “Italy is boot-shaped,” and got away with it, then my utterance is true enough. The standards of precision are sufficiently relaxed. But if you have just denied that Italy is boot-shaped, and carefully pointed out the differences, then my utterance is far from true enough—the standards of precision are too exacting. The acceptability of what I say here depends on the conversational score, on the context, which in turn depends on what has been said before. The extension of “hexagonal” shifts with changes of context. Or, for another example, suppose I say “The pavement is flat” under standards of flatness where the bumps in the pavement are too small to be relevant. Then what I say is true. But if the conversational score changes, and I say “The pavement is flat” under raised standards of flatness, what I say will no longer be true. But “[t]hat does not alter the fact that it *was* true enough *in its original context*.”⁹ Like the extension of “hexagonal,” the extension of “flat” changes with the context.

According to Stalnaker, contexts may be identified with the shared presuppositions of conversants; according to Lewis, the shared presuppositions are a component of the context. From their accounts of context we can extract the idea of tracking context-change by keeping a running record of shifts in the salient information presumed to be available to the participants. A number of semanticists and linguists have also developed this idea. It can be found for example in Irene Heim’s file change semantics,¹⁰ Grosz and Sidner’s dynamic theory of discourse structure,¹¹ and the literature on the distinction between “given” and “new” information.¹²

29.4 A CONTEXTUAL ANALYSIS OF REVENGE

Contextualists about truth, of both types, can draw on this conception of context-change to provide a plausible account of the revenge discourse. For simplicity, I shall illustrate the application of this literature to the revenge liar in terms of the first type of theory, where “true” is taken to be context-sensitive. But the second type of theory can equally well draw on this account of context and context-change, as we will see.

Let’s return to our revenge discourse. There are two initial segments: (1) the production of the revenge sentence R, and (2) the reasoning to the conclusion that R is semantically defective. Turning first to *repetition*, we add two further segments: (3) the

⁹ Lewis (1979: 246).

¹⁰ See Heim (1988). A “file” contains all the information that has been conveyed up to that point—and the file is continually updated as the discourse moves on. Heim’s notion of common ground is more fine-grained than Stalnaker’s, since Heim’s account is more sensitive to the subsentential structure of sentences.

¹¹ Grosz and Sidner (1986).

¹² See e.g. Halliday (1967); Halliday and Hasan (1976); Chafe (1976); Clark and Havilland (1977); Clark and Clark (1977); Allerton (1978); Prince (1981); Brown and Yule (1983).

inference to the sentence S, and (4) the conclusion that S is true. Consider the transition from the second to the third segment of the discourse. The culmination of the reasoning of the second segment is the proposition that R is semantically defective. This is new, salient information that becomes part of the common ground. So in the transition from the second segment to the third, there is a context change—a shift in the body of information that is presumed to be available. According to Heim's account, for example, I will register this shift by updating the file card that stores information about R: I will now add the entry "is semantically defective." Grosz and Sidner's focusing structure distinguishes the salient objects, properties and relations at each point of the discourse—and as we move from the second segment to the third, it will distinguish the defectiveness of R.

Let us say that the new context associated with the third segment is *reflective with respect to R*. In general, a context associated with a given point of a discourse is *reflective with respect to a given expression* if at that point it is part of the common ground that the expression is semantically defective or pathological. So as we move from the second segment to the third, there is a context-change—a shift to a context that is reflective with respect to R. This context-change is an essential ingredient of revenge reasoning. It is the mark of such reasoning that we shift to a context which is reflective with respect to a semantically defective expression.

Thus far, we have seen that content acts on context—newly available information changes the context. But there is a two-way interaction between context and content: context also acts on content. Recall the challenge posed by *repetition*: (R) and (S) are tokens of the same type with the same linguistic meaning, yet one is semantically defective and the other is true. The task is to explain how the changes in context produce this phenomenon. If context acts on content, we would expect there to be an expression in the discourse that is sensitive to context-change. On the present contextual account, "true" is that expression. (Looking at the terms that appear in the revenge discourse, it is very hard to see a viable alternative.)

Let c_R be the initial context associated with the first segment in which R is produced. The first use of "true" is a component of the sentence R, and so occurs in the context c_R . So we represent this first use of "true" by " true_{c_R} ," so that R is represented by:

(R) The sentence written at the third line of Section 2 of this article is not true_{c_R} .

This representation does not commit us to the claim that "true" is context-sensitive—we are only marking the fact that this use of "true" occurs in context c_R . We will continue to attach the subscript " c_R " to each subsequent use of "true" if there is no change of extension. (If context has no effect on the extension of "true"—if "true" is a predicate constant—then the continued appearance of the subscript " c_R " is vacuous and will indicate no more than this: every use of "true" is coextensive with every other. If, on the other hand, "true" is context-sensitive, then the subscript c_R will reappear only if subsequent uses of "true" inherit the extension that the context c_R determined for this first use of "true.")

Now R assesses a certain sentence as not true_{cR}. Under what conditions is a sentence true_{cR} or not? That's determined by the truth_{cR}-schema:

s is true_{cR} iff p

where s is an expression that refers to the sentence p. Now, given the empirical circumstances, R assesses itself. So we apply the truth_{cR}-schema to R.¹³ Suppose R is true_{cR} (i.e. we suppose the left-hand side of the schema)—then, given the schema, it follows that the sentence written at the second line of Section 2 of this article is not true_{cR} (i.e. we infer the right-hand side). So we reach a contradiction. Suppose that R is false_{cR}—then the sentence written at the second line of Section 2 of this article is not true_{cR}, and, by the truth_{cR}-schema, it follows that (R) is true_{cR}, and we have a contradiction again. This is the reasoning at the second segment of the revenge discourse: when we assess R by the truth_{cR}-schema, we find that it's semantically defective, since it cannot be given truth_{cR} conditions.

Now when we infer that R is not true, and thereby *repeat* R, our use of "true" in the third segment is again to be represented by "true_{cR}"—we infer that R is not true_{cR}, since R is neither true_{cR} nor false_{cR}, and cannot be evaluated by the truth_{cR}-schema. So S is to be represented this way:

(S) The sentence written at the third line of Section 2 of this article is not true_{cR}.

In a strict sense, then, S does indeed repeat R—it is composed of the same words with the same meanings *and* the same extensions. But here R is repeated in a context that is *reflective with respect to R*—i.e. where it is recognized that R is semantically defective. And R is defective because it cannot be evaluated by its associated schema, namely the c_R-schema.

So at the third segment, the effect of the shift of context is to break the link between R and the truth_{cR}-schema—and S announces this separation.¹⁴ The truth_{cR}-schema is abandoned as an evaluating schema for R. And since S is an exact repetition of R, the truth_{cR}-schema cannot serve as an evaluating schema for S either. To anticipate the fourth segment, we are able to conclude that S is true just because we have abandoned the truth_{cR}-schema.

¹³ Burge identifies a Gricean pragmatic implicature here. Suppose E is an *evaluating* sentence, a sentence that applies "true" to some sentence or class of sentences (as in "Socrates' last utterance was true" or "Nothing Smith says is true"). And let the *associated truth-schema* for E be the schema in which "true" is co-extensive with the occurrence of "true" in E. Then Burge claims that there is the following pragmatic implicature: the sentence or class of sentences to which E refers or which it quantifies over is to be evaluated by the associated truth-schema for E. In the particular case of R, R is to be evaluated by the truth_{cR}-schema. For a critical discussion of the role of implicature in Burge's theory, see Gaifman (1992: 259–60).

¹⁴ Burge puts this in Gricean terms, as the cancellation of an implicature. The implicature or background assumption that R is to be assessed by the truth_{cR}-schema is canceled.

At the fourth segment we conclude that *S* is true. This occurrence of “true” does not inherit its extension from earlier occurrences—unlike the occurrence of “true” in *S*. Here we should keep in mind that we’ve come to a true conclusion through valid reasoning. According to our conclusion, *S* is true. However, *S* is not true_{c_R} —if we assess *S* by the truth_{c_R} -schema, we obtain a contradiction, just as we did with *R*. But our conclusion is not contradictory—it is true. So here, there is a shift in extension—*S* is not true_{c_R} , but *S* is true_{r_R} , let us say, where the subscript “ r_R ” stands for the context associated with the third and fourth segments, a context reflective with respect to *R*. The occurrences of “true” represented by “ true_{c_R} ” and “ true_{r_R} ” have different extensions: *S* is in the extension of “ true_{r_R} ” but not in the extension of “ true_{c_R} ”.

What produces this shift in the extension of “true”? The change in context—specifically, the shift to a context which is reflective with respect to *R*. At the third stage, the reflective character of the context had the effect of disengaging *R* from the c_R -schema. Now, at the fourth stage, it has the effect of engaging a new schema—the reflective r_R -schema. When we assess *S*, and declare that it is true, we assess it in a context where it is part of the common ground that *R* is pathological. The schema by which we assess *S* provides an assessment of *S* in the light of *R*’s pathologicity. Here’s what the instance of the r_R -schema looks like:

S is true_{r_R} iff the sentence written at the third line of Section 2 of this article is not true_{c_R} .

Given the information that forms part of the common ground at the third and fourth segments—that *R* is defective and is not true (i.e. not true_{c_R})—the right-hand side of the biconditional holds, and so we infer the left-hand side. And that is what is going on in the fourth segment.

We’ve seen that with the shift to a reflective context, *R* is disengaged from the truth_{c_R} -schema—this is announced by the repetition *S*, according to which *R* is not true_{c_R} . So even though “true” in *S* occurs in the reflective context r_R , this occurrence is not represented by “ true_{r_R} .” Call an indexical *narrow* if its extension is supplied entirely by the context of utterance. Indexicals such as “I,” “here,” and “now” are often taken to be narrow in this sense (supply the speaker in the context, and the extension of “I” is fully determined, and analogously for “here” and “now”). Jason Stanley has taken contextualists about truth to be claiming that “true” is a narrow indexical, and he argues that this claim is put into doubt by the observation that the vast number of cases of unobvious context-dependence do not involve narrow indexicality.¹⁵ But according to the present contextual line, “true” is not a narrow indexical. An occurrence of “true” in the context r_R is not automatically to be represented by “ true_{r_R} ”—the extension of an occurrence of “true” in a reflective context may be inherited from an earlier context, as with “true” in the repetition *S*.¹⁶

¹⁵ See Stanley (2000: 430–1).

¹⁶ Stanley contrasts narrow indexicals with e.g. the comparative adjective “small”—the sentence “Most species have members that are small” has a reading (“Most species *S* have members that are small for *S*”) where the extension of “small” is not determined by the standard made salient by the context of

In a nutshell, we explain our different assessments of R and S this way: we assess R by the unreflective c_R -schema, and we assess S by the reflective r_R -schema. With the change in context, there is a change in the evaluating schema. There is no intrinsic difference between R and S—the difference lies in the schema by which they are assessed.

The analysis of *Repetition* extends naturally to *Rehabilitation*. On the present contextual analysis, R and S do say exactly the same thing, and so we conclude that R, like S, is true (i.e. true_{rR}), where R, like S, is assessed by the reflective truth $_{rR}$ -schema. We evaluate R on the basis of its defectiveness—R is defective, and so not true (i.e. not true_{cR}), and it says that it's not true_{cR} , and so we evaluate it as true (i.e. true_{rR}). Compare Lewis's treatment of "hexagonal" or "flat." Sometimes an utterance of "France is hexagonal" (or "The pavement is flat") is true, and sometimes it isn't. The extension of the predicates "hexagonal" or "flat" depend on the conversational score, in particular on the standards of assessment that are in force. In an analogous way, whether or not it is true to say that R is true will depend on the standard of assessment: do we apply the unreflective c_R -schema, or the reflective r_R -schema?

And now we can also give an account of *Oscillation*. The oscillation between opposing values for R is analyzed as the oscillation between the two sides of the reflective truth $_{rR}$ -schema for R:

R is true_{rR} iff the sentence written at the second line of Section 2 of this article is not true_{cR} .

R is true_{rR} just in case R is not true_{cR} —but there is no contradiction. Indeed, (R) is true_{rR} because R is not true_{cR} , and that's what it says.

When we use "true" in an evaluation of R, the extension of the truth predicate depends on whether or not the given context is reflective with respect to R. We have identified a contextual parameter—the *reflective status* of a context—to which the term "true" is sensitive. If we do not attend to revenge reasoning, the claim that "true" is a context-sensitive term may come as a surprise, and *reflective status* will not be an obvious contextual coordinate (unlike the familiar coordinates of speaker, time, and place, for example). But on the present contextual line, once we pay careful attention to revenge discourses where we reason past pathology, it seems natural and intuitive to conclude that "true" is indeed sensitive to the reflective status of a context. Cresswell wrote,

It seems to me impossible to lay down in advance what sort of thing is going to count [as a relevant feature of context] . . . The moral here seems to be that there is no way of specifying a finite list of contextual coordinates.¹⁷

utterance. One might wonder, given that "true" presents such a special case, why the unobvious context-dependence of "true" should go the way of other cases of unobvious context-dependence. But anyway, the present contextual account of "true" does not claim that "true" is a narrow indexical. As with "small," there are uses of "true" whose extensions are not supplied entirely by the standard made salient by the context of utterance.

¹⁷ Cresswell (1972), quoted in Lewis (1980: 30). One target of Cresswell's remark is Lewis (1970), and Lewis takes Cresswell's criticism to heart in Lewis (1980). In a somewhat similar vein, Kaplan

Along with Cresswell, Lewis, Stalnaker, and others, we should be open to contextual coordinates beyond the familiar ones. If we recognize that content acts on context, that new information or new presuppositions can change the context, then we can identify contextual coordinates that we might otherwise miss. Reflective status is such a coordinate.

29.5 TRUTH AND HIERARCHY

The first type of contextualist theory claims that “true” is a context-sensitive predicate—in the course of the revenge reasoning, there is a shift in the extension of “true.” What is the relation between the occurrences of “true” that we have distinguished as “true_{c_R}” and “true_{r_R}”? According to Burge, these occurrences correspond to distinct levels in a Tarskian hierarchy (Burge 1979). This contextual-hierarchical account is endorsed in Koons 1992.¹⁸ We can think of “true” as it occurs initially in R as indexed to a certain level represented by the number *i*.¹⁹ So now we represent the initial occurrence of “true” as “true_{*i*},” replacing the contextual subscript “c_R” by the numerical subscript that indicates the level fixed by the context. And when we “rehabilitate” R, and declare it true, we do so at a higher level than *i*, say *k*: (R) is not true_{*i*}, but it is true_{*k*}, where *k* > *i*.

According to Burge, the level of a sentence is established in context by certain material principles of interpretation: Justice, Verity, Beauty (or Minimalization).²⁰ To motivate Justice, consider for instance a loop, where Plato and Aristotle each say of the other’s utterance at time *t* that it isn’t true. It would be arbitrary to treat these utterances differently—both should count as pathological, and each should be assigned the same level. According to Justice, then, “one should not give one statement truth conditions instead of another without some reason.”²¹ According to Verity, subscripts on “true” should be assigned “so as to maximize the interpreter’s ability to give a sentence truth conditions by way of a truth schema” (Martin 1984: 109). So, for example, if I say “Everything Descartes said that does not concern mechanics is true,” then the subscript on “true” should be high enough to give truth conditions to everything Descartes said that does not concern mechanics (where the level may be quite unknown to me or an interpreter of my utterance). Descartes might have evaluated utterances that themselves contain

writes: “context *provides* whatever parameters are needed” (Kaplan 1989: 591), though Kaplan’s remark is restricted to expressions that are “directly referential.”

¹⁸ In Koons (1992: ch. 6), Koons argues that the hierarchical theories of Burge, Barwise and Etchemendy, and Gaifman are special cases of a more general theory, and then, in ch. 7, applies this theory to doxic paradoxes. For further discussion of Koons, see e.g. Juhl (1997).

¹⁹ In his presentation of the formal theory, Burge explicitly accommodates only finite levels. As he acknowledges in the postscript to Burge (1979), provisions would need to be made for extending the constructions into the transfinite (see Martin 1984: 115).

²⁰ This is the ordering that Burge suggests in the postscript to Burge (1979) in Martin (1984: 115).

²¹ Burge in Martin (1984: 110).

the truth predicate; if so, and if Descartes's evaluations are not looped with others, or in some way implicated in paradox, then my utterance will be at a level higher than any of Descartes's utterances, so as to confer truth conditions on Descartes's non-mechanical statements. By Verity, we need only worry about paradox if it is forced upon us—as it is with the sentence *R*, since *R* itself satisfies the definite description that occurs in *R*. While Verity requires the level to be high enough to fit the interpreter's purposes, Beauty requires the level to be the lowest compatible with Justice and Verity. Burge's main motivation for Beauty is that it simplifies his formal account (Burge 1982: 359). The application of these pragmatic principles allows truth to find its own level—the level is not merely a product of the speaker's or hearer's intentions, but also of facts about the context of use and general conventions.²²

In his formal account, Burge offers three alternative constructions, all of which aim to define the notion of a pathological_{*i*} sentence (i.e. a sentence that fails to have truth_{*i*} conditions). On the first construction—an analogue of Tarski's construction—the pathological_{*i*} sentences are all and only those sentences that contain “true_{*k*},” for $1 \leq i \leq k$.²³ (To put things the other way around, a sentence has truth_{*i*} conditions only if any occurrence of “true” it contains is of a level lower than *i*.) And the truth-schema takes a conditional form:

If ‘ φ ’ is not pathological_{*i*}, then ‘ φ ’ is true_{*i*} iff φ .

Burge's second construction liberalizes the definition of a pathological_{*i*} sentence. This construction accommodates the intuition that if the sentence “Snow is white” is true_{*i*}, then the sentence “Snow is white or *a* is true_{*k*}” (where “*a*” denotes a sentence) should also count as true_{*i*}, even where $i \leq k$.²⁴ In general, on the second construction, any logically valid_{*i*} inference from true_{*i*} sentences counts as true_{*i*}. Burge's third construction is still more liberal: it counts as true_{*i*} not only logically valid_{*i*} inferences from true_{*i*} sentences, but also any sentence that says of a true_{*i*} sentence that it is true_{*i*}.²⁵ Though the second and third constructions are more liberal than the first, at the heart of all three is the application of Tarski's levels-of-language idea to natural language.

²² See Burge (1979). In this way, Burge escapes Kripke's criticism of theories (perhaps like that of Parsons 1974) where level is explicitly or implicitly assigned in advance by speakers (see Kripke 1975: 58–60).

²³ The first construction departs from the Tarskian hierarchy in allowing natural-language sentences that lead to paradox to be well-formed. And sentences that Tarski would count ungrammatical can be given truth conditions at appropriate higher levels.

²⁴ The intuition here also motivates Kleene's strong tables (see Kleene 1952: section 64, 332–40).

²⁵ The intuition behind the third construction is that semantical evaluation should be grounded, in a sense like that of Kripke (Kripke 1975). Intuitively, the evaluation

(E) “Snow is white” is true

is grounded, because it evaluates a nonsemantical sentence, and the evaluation “*E* is true” is also grounded, because it evaluates an evaluation which is grounded, and so on. On Burge's third construction, the true_{*i*} sentences will include those sentences that say of true_{*i*} sentences that they

Burge goes on to show how his account can handle cases thought by some to be problematic for Tarskian approaches. Here is an example adapted from Kripke.²⁶ Suppose Dean asserts,

- (1) All of Nixon's utterances about Watergate are untrue,

while Nixon asserts,

- (2) Everything Dean says about Watergate is untrue.

This loop poses a problem for an orthodox "levels" approach, where a statement of a given level can speak only of the truth or falsity of statements of lower levels. It seems that Dean intends (1) to include Nixon's utterance within its scope—and Nixon likewise intends (2) to include Dean's utterance in its scope. But suppose (1) and (2) are assigned the same level—then neither can include the other in its scope. And if (1) is assigned a higher level than (2), then (2) cannot include (1) in its scope; and vice versa. But, as Kripke points out (1975: 60), it is perfectly possible for (1) and (2) to have unambiguous truth values. For example, suppose Dean has made at least one true statement about Watergate. Then (2) is untrue. And if everything else that Nixon says about Watergate is untrue as well, then (1) is true. These intuitive assignments require each statement to include the other in its scope.

Burge resolves the case as follows.²⁷ By Justice, "true" in (1) and (2) receive the same subscript, say *i*. By Verity, *i* is high enough to allow application of the truth_{*i*} schema to any utterance of Dean's or Nixon's other than (1) or (2), and, by Beauty, no higher. The level is fixed by the pragmatic principles and the empirical facts—the level need not be known by anyone. On the third construction (Burge's preferred account in Burge 1982), the results are as follows. If Dean made at least one true statement about Watergate (by Verity, this will be a true_{*i*} statement), (2) is untrue_{*i*}. (This is because the negation of (2) is true_{*i*}, since it is a logically valid_{*i*} inference from a true_{*i*} sentence.) And if everything else that Nixon says about Watergate is untrue (i.e. untrue_{*i*}, by Verity), then (1) is true_{*i*}. These are the intuitive assignments. Now the empirical circumstances may instead conspire to

are true_{*i*}. Consider e.g. the true_{*i*} sentences. The true_{*i*} sentences are obtained as follows: "... take all nonsemantical true_{*i*} sentences, add sentences logically derivable from them; add all sentences that say that these sentences are true_{*i*} (and that their negations are not true_{*i*}); then add sentences logically derivable from them; add all sentences that say that these sentences are true, (and that their negations are not true_{*i*}); and so on" (Burge 1979:104). Then we do the same for "true_{*j*}," starting with all true_{*j*} sentences that are either nonsemantical or contain only "true_{*i*}." And so on, through the levels. The third construction departs from Kripke's notion of groundedness in the following respect: evaluations can be grounded (or "rooted," to use Burge's preferred term) in lower-level semantical evaluations, and need not be traceable back to a nonsemantical sentence; e.g. the evaluation that (R) is true_{*k*} is grounded (rooted) in the evaluation that (R) is not true_{*i*} (where $i \leq k$).

²⁶ Kripke (1975: 59–60).

²⁷ Burge (1979: 110–11; 1982: 360–1).

make (1) and (2) defective—suppose that all of Dean’s utterances about Watergate other than (1), and all of Nixon’s other than (2), are untrue_i. Then (1) and (2) are caught in a loop—each depends on the other for its truth value. Then each is pathological_i—they have no truth_i conditions and so are not true_i. But that’s what each says about the other—so both are true_{i+1}. At this higher level of language, where (1) and (2) are evaluated in the light of their pathology, the statements are evaluated as true. This seems to be the intuitively correct result, and runs parallel to the intuitions associated with the revenge liar.

29.6 TRUTH AND SINGULARITIES

Like Burge’s theory, the singularity theory takes “true” to be a context-sensitive predicate.²⁸ But while Burge endorses a Tarskian hierarchy for natural language, the singularity theory does not stratify truth into levels. Rather, a particular use of “true” is minimally restricted, applying globally except for certain “singularities,” where its application breaks down. These singularities of “true” vary with the context.

The singularity theory is in the spirit of remarks of Gödel’s. Gödel notes that Russell’s theory brings in a new idea for the solution of the paradoxes:

It consists in blaming the paradoxes . . . on the assumption that every concept gives a meaningful proposition, if asserted for any arbitrary object or objects as arguments.²⁹

Gödel goes on to say that the simple theory of types carries through this idea on the basis of a further restrictive principle, by which objects are grouped into mutually exclusive ranges of significance, or types, arranged in a hierarchy. Gödel suggests that we reject this principle, while retaining the idea that not every concept gives a meaningful proposition for any object as argument:

It is not impossible that the idea of limited ranges of significance could be carried out without the above restrictive principle. It might even turn out that it is possible to assume every concept to be significant everywhere except for certain “singular points” or “limiting points,” so that the paradoxes would appear as something analogous to dividing by zero. Such a system would be most satisfying in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e. they could then be considered to give an essentially correct, only somewhat “blurred,” picture of the real state of affairs.³⁰

The singularity theory applies this tantalizing idea to truth, suggesting that a use of “true” in natural language has singularities determined by the context.

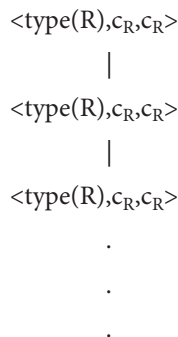
²⁸ See Simmons (1993; 2018).

²⁹ Kurt Gödel (1944) in Schilpp (1944: 228).

³⁰ Kurt Gödel (1944) in Schilpp (1944: 229).

Consider again the liar sentence R. According to the contextual analysis of section 29.4, the occurrence of “true” in R is represented by “true_{c_R}”, and the associated evaluating schema is the truth_{c_R}-schema. So we can represent R by the ordered triple $\langle \text{type}(R), c_R, c_R \rangle$, where the first member of the triple indicates the type of R, the second indicates that the occurrence of “true” in R is represented by “true_{c_R}”, and the third indicates that the associated schema is the truth_{c_R}-schema. The repetition S may be represented by $\langle \text{type}(R), c_R, r_R \rangle$, since $\text{type}(S) = \text{type}(R)$. The triples for R and S have the same first two members—S is a repetition of R. But R and S have different associated schemas. The schema associated with S is the truth_{r_R}-schema, which is reflective with respect to R. The only difference between R and S is the schema by which they are assessed in the course of the revenge reasoning. Call the representation of R by $\langle \text{type}(R), c_R, c_R \rangle$ the *primary representation* of R, because the third member of the triple indicates R’s associated schema—R can be assessed by other schemas, but we are primarily interested in the schema by which it is assessed in the revenge reasoning. Similarly, the primary representation of S is $\langle \text{type}(R), c_R, r_R \rangle$.

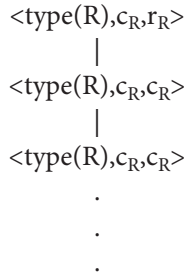
The main task of the singularity theory is to identify singularities of a given occurrence of “true.” We understand by a *sentence* a sentence type in a context. Let the *determination set* of a given sentence σ be the set of those sentences to which σ makes reference. In the case of R, this is the unit set containing just R. The *primary tree* for σ is constructed as follows. At the top is the primary representation of σ . At the second tier are the primary representations of the members of σ ’s determination set. At the third tier are the primary representations of the members of the determination sets of sentences represented at the second tier. And so on. The primary tree for R looks like this:



This is a (single-branched) tree with an infinite branch, indicating semantic pathology: R cannot be assessed by its associated schema, the truth_{c_R}-schema. In general, if the primary representation $\langle \text{type}(\sigma), c_\alpha, c_\beta \rangle$ of a sentence σ repeats on an infinite branch of σ ’s primary tree, then σ is pathological, and a *singularity* of “true_{c_{\beta}}.” In particular, R is pathological, and a singularity of “true_{c_R}.” According to the singularity theory, if σ is a singularity of “truth_{c_{\beta}},” then σ is excluded from the extension of “true_{c_{\beta}}.” So R is

excluded from the extension of “true_{cR}.” So R isn’t true_{cR}, just as R says. And this explains *Rehabilitation*: once we reflect on R’s pathology, and exclude the singularity R from the extension of true_{cR}, we evaluate R as true upon reflection—i.e. true_{rR}.

The primary tree for S is:



The primary representation of S does not repeat on this infinite branch, so S is not pathological. Unlike R, S *can* be assessed by its associated schema: S is true_{rR}. This explains *Repetition*.

In the case where Nixon’s (1) and Dean’s (2) produce pathology, their primary trees—which will have many branches, given the size of the determination sets—will each have an infinite branch which cycles continuously through (1) and (2).³¹ These branches indicate that (1) and (2) are pathological, and that (2) is a singularity of “true” in (1), and (1) is a singularity of (2). Since these singularities must be excluded, (1) and (2) are to be both be reflectively evaluated as true.³²

A guiding principle of the singularity theory is *Minimality*: restrictions on occurrences of “true” are kept to a minimum. We are to restrict the application of “true” only when there is reason to do so. Suppose you say, “‘Snow is white’ is true.” Here your use of “true” is quite unproblematic. Should R be excluded from its extension? Minimality says not. R *isn’t* true_{cR}, and that’s what it says—and that’s why it’s true_{rR}. And for the same reason, R can be counted as true in your neutral context of utterance—we have no reason to tie your innocent utterance to R’s pathology. By Minimality, if we can count R as true in a given context of your utterance, then we must so count it. Minimality treats your use of “true” as reflective with respect to R—though not, of course, as explicitly so.³³ In general,

³¹ In different empirical circumstances, as we saw, (1) and (2) may receive determinate truth values. In Simmons (1993), these cases are treated in accordance with Kleene’s strong tables via a formal account of grounding, and the notion of a pruned tree. Intuitively, the idea is that in cases where definite truth values can be assigned, appropriate pruning removes any infinite branches—so pathology is not indicated.

³² A full account of this case along singularity lines can be found in Simmons (1993: 142–5).

³³ For another example of Minimality at work, suppose Max and Claire say the following:

(M) What Claire is saying is true.

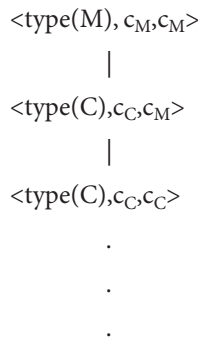
(C): What Claire is saying is not true. (cont.)

given an occurrence of “true,” we exclude from its extension only those sentences we have to—its singularities (and every occurrence of “true” will have singularities).³⁴ The Tarskian truth schema is minimally restricted:

For any context α , if $\ulcorner \sigma \urcorner$ is not a singularity of “true $_{\alpha}$,” then $\ulcorner \sigma \urcorner$ is true $_{\alpha}$ iff σ .

It can be argued that by adopting Minimality, we respect a basic intuition about predicates: intuitively, we take a predicate to pick out everything with the property that the predicate denotes. According to the singularity account, the scope of “true” on any occasion of use is as close to global as it can be.³⁵

The primary tree for Claire’s utterance indicates that C is a singularity of “true” in C: the triple $\langle \text{type}(C), c_C, c_C \rangle$ repeats on an infinite branch. So C is not true $_{c_C}$, which is what C says. So C is true in any context reflective with respect to C. The primary tree for M looks like this:



Notice that the second member of the triple at the first tier becomes the third member of the triple at the second tier—C is assessed via the schema associated with the occurrence of “true” in M, since M is evaluating C. Similarly, the second member of the triple at the second tier becomes the third member of the triple at the third tier (since C is evaluating C). The primary representation of M does not repeat on this infinite branch, so M is not pathological—intuitively, M stands above the loop in which C is caught. *In accordance with Minimality, the context c_M is treated as reflective with respect to C.* Since C is true in any context reflective with respect to C, Minimality requires that C is true $_{c_M}$ —and so given what M says, M is true $_{c_M}$.

³⁴ Anaphoric liars show that every occurrence of “true” has singularities. We can produce an anaphoric liar sentence by extending your innocent utterance: “‘Snow is white’ is true but this very sentence isn’t.” An anaphoric liar can help us see the anti-hierarchical nature of the singularity account. Suppose we reflect on R and declare, “R is true.” The occurrence of “true” here occurs in a context reflective with respect to R—R is not a singularity of this occurrence. But this occurrence of “true” also has its own singularities—suppose we append “but this very sentence isn’t.” But by Minimality, this anaphoric liar is not a singularity of the occurrence of “true” in R. The two occurrences of “true” have different singularities. We do not move from a less comprehensive truth predicate to a more comprehensive one. Neither extension includes the other.

³⁵ For further discussion of the singularity theory, see e.g. Grim (1995); Antonelli (1996); Hardy (1997); Beall (2003); Gauker (2006). For the extension of the singularity theory to paradoxes of denotation and Russell’s paradox, see Simmons (1994; 2000; 2018).

29.7 CONTEXT AND QUANTIFIER DOMAINS

Both Burge's Tarskian theory and the singularity theory locate the context-sensitivity associated with the liar in the truth predicate—"true" shifts its extension according to context. Another major contextualist approach locates the context-sensitivity elsewhere, in a more general setting not limited to truth. If I say "There's no beer left," I do not mean there is no beer left in the entire world—I mean there's none left in the refrigerator. Context determines the domain of quantification here. The idea that the context-dependence of truth is derived from the context-dependence of quantifier domains was first suggested by Parsons and developed in a fully rigorous way by Glanzberg.

Parsons suggests that we can accommodate the conclusion of our revenge reasoning, that the liar sentence is true, if we assume that this evaluation "presupposes a more comprehensive scheme of interpretation than the discourse up to that point" (Parsons 1974: 35). The final evaluation "involves a semantical reflection that could be viewed as involving taking into one's ontology a proposition that had not been admitted before" (ibid.). Glanzberg makes these ideas precise in Glanzberg (2001; 2004).

Glanzberg takes *propositions* to be the bearers of truth, preferring propositions to the two main alternatives, sentence-context pairs, and utterances.³⁶ Sentences are true only in a derivative sense: a sentence is true iff it expresses a true proposition. So we should consider the liar sentence P, introduced above:

(P) P does not express a true proposition.

This yields a revenge liar in terms of propositions. The reasoning that Glanzberg lays out rests on three natural assumptions.³⁷ First, a sentence expresses a unique proposition—this is uncontroversial, since, in deriving the paradoxical conclusion, Glanzberg assumes that context plays no role of any kind (and ambiguity is set aside). Second, truth is an extensional property of propositions (if propositions q and r are identical, then q is true iff r is true). The third assumption connects truth and the expression relation:

(T-Exp) If the sentence $\ulcorner \sigma \urcorner$ expresses the proposition q, then q is true iff σ .

The revenge liar reasoning runs as follows:³⁸

- (1) Suppose that the sentence P expresses a proposition p.

³⁶ See Glanzberg (2001: 226–7).

³⁷ See Glanzberg (2004: 33–4; also 2001: 229).

³⁸ For the sake of simplicity, the applications of the first two assumptions are not made explicit here. See Glanzberg (2004: 33–4) for details.

(2) Suppose *p* is true.

Then, (3) *P* does not express a true proposition (from (2), by T-Exp).

So, (4) *p* is not true (from 3).

(5) Suppose *p* is not true.

Then, (6) *P* expresses a true proposition (by T-Exp).

So, (7) *p* is true (from 6).

From lines (2)–(7), we have established that *p* is true iff *p* is not true, under the assumption at (1). So we have a *reductio* proof that the supposition at (1) is false. That is, we conclude:

(8) *P* does not express a proposition.

But we need not stop here. It follows from (8) that

(9) *P* does not express a true proposition.

But this is just P again. (This may be regarded as a propositional version of *Repetition*.) So we have proved *P*—and so *P* is true. (This is *Rehabilitation*.) But for a sentence to be true is just for it to express a true proposition. It follows that

(10) (*P*) does express a proposition.

(8) and (10) are contradictory, and we are landed in paradox.

Glanzberg focuses on (8) and (9). Both are true, because both are established by sound reasoning. But since (8) is true, there is no proposition expressed by (*P*). But for (9)—i.e. *P*—to be true, there is a proposition expressed by *P*. We *proved* that *P* cannot express a proposition, and then we *proved* that it can. The proofs rest on the three assumptions, but these assumptions seem beyond any suspicion.³⁹ But how can it be that *P* at first fails to express a proposition, and then succeeds? Without admitting the context-dependence of the liar sentence, the question seems unanswerable. According to Glanzberg, we are forced to conclude that *P* exhibits some context-dependence. Since propositions are the truth-bearers, and the truth values of propositions do not vary from context to context, the predicate “true” will not itself be context-dependent. But things are different with the expression relation: it is perfectly possible for a sentence to express a proposition in one context but not in another. So we have a way of answering the question: in the context of (8), there is no proposition expressed by the liar sentence *P*, but in the context of (9), there is. There is a shift in the domain of the propositional quantifier. The domain

³⁹ Perhaps the sentential analogue of (T-Exp) can be challenged—but (T-Exp) is in terms of propositions, and it seems much harder to deny.

of propositions associated with the context of (8) does not contain a proposition for P to express; the expanded domain associated with the context of (9) does. The challenge now is to show how this expansion fits with standard ideas from linguistics and philosophy of language.

To meet this challenge, Glanzberg makes use of the idea articulated in section 29.3, that context provides a running record of information available at a given point in a discourse. In particular, he draws on the extensive work in the literature on the notions of *salience* and *topic*, in order to make it clear that taking context to include a running list of salient items is well-motivated, and quite independent of the liar.⁴⁰ These salient items may be individuals, relations, or domains.⁴¹ Glanzberg calls this component of context the *salience structure*.

Consider now the contexts of (8) and (9). The context of (8) involves an assertion of “(P) does not express a proposition.” This is the first point in the proof where there are no undischarged premises (at this point, we complete the *reductio* argument that we started at (1)). So this is the first point in the proof where an unconditional assertion involving the expression relation is produced. Consequently, Glanzberg argues, it is here that the expression relation is accepted as *salient* in the discourse—and there is an expansion of the salience structure to include the expression relation. So between (8) and (9) there is a genuine difference in context.

Glanzberg goes on to show how this expansion of the salience structure can lead to an expansion of the domain of truth conditions, so as to allow the expression of a proposition that could not previously be expressed. Glanzberg begins with the familiar idea that when we express a proposition in a context, we divide the class of possible worlds into the class of those in which the claim is true, and those in which it is not true. Truth conditions are understood as possible worlds, and a proposition is identified with the class of those possible worlds (truth conditions) in which it is true. But now speakers do not, of course, survey an independently given domain of worlds, and decide which are in the proposition and which are not. Rather, they have to rely on linguistic resources to identify worlds and decide which are in and which are out. So the worlds, or truth conditions, associated with propositions should not go beyond those that can be differentiated by linguistic means. According to Glanzberg, then, the domain of truth conditions (worlds) is constrained by the resources speakers have for expressing propositions. And these resources may shift as the context shifts. So the context-dependence associated with the liar is of an *extraordinary* kind—as Glanzberg puts it, “it is the dependence of the background domain of truth conditions upon context”

⁴⁰ For the general idea that context is a running record of information, Glanzberg cites Stalnaker (1978) and Lewis (1979). For work on context and salience, Glanzberg cites Karttunen (1976); Heim (1988); Kamp (1984); for work on the role of topic in context, Büring (1999); Portner and Yabushita (1998); Reinhart (1981); Roberts (1996); Vallduvi (1992); van Kuppevelt (1995); and von Fintel (1994); and for similar ideas applied to natural language processing, the focus spaces of Grosz (1977), and the focusing structures of Grosz and Sidner (1986).

⁴¹ Further linguistic support for this claim is provided in Glanzberg (2002).

(2004: 19). Where we have *ordinary* context-dependence, the truth conditions of a sentence can change with a change in context—but there is no shift in the domain of truth conditions itself. But the context-dependence exhibited by the liar is not of an ordinary kind: the domain of truth conditions *itself* shifts with context.⁴²

It remains to be shown that the domain of truth conditions can not only change with a shift in context, but can *expand* with a shift in context. The relevant shift in context is the expansion of the salience structure to accommodate the *expression* relation. Glanzberg points out that if speakers have available semantic predicates such as “express” or “true,” they can individuate very complicated truth conditions. Using “true,” speakers can for example accommodate truth conditions associated with sentences of potentially infinite length. Consider the sentence “Everything the Pope says is true”—this is equivalent to an infinite conjunction of conditionals (of the form “If the Pope says ‘s’, then s”), and it has truth conditions of enormous complexity. So when the semantic relation of expression is added to the salience structure, it is to be expected that the domain of truth conditions will expand, given that this domain is constrained by the resources that speakers have available in a context.

This is a sketch of a solution to what Glanzberg calls “the expansion problem,” and in Glanzberg (2004), he goes on to provide a formal model of extraordinary context-dependence and expansion. The main components are: a formal treatment of salience structures as contexts; the construction of a domain of truth conditions relative to a salience structure; and a proof that as the context changes in the course of the revenge discourse, this domain of truth conditions expands.⁴³ This provides a formal underpinning for a resolution of the liar. We were confronted by apparently paradoxical conclusions at (8) and (9): according to (8), the liar sentence does not express a proposition, but (9) requires that it does. But both (8) and (9) are in fact true: relative to the initial context, with a smaller domain of truth conditions, there is no proposition for P to express. Relative to the expanded domain of the new context, there is.⁴⁴

⁴² This “extraordinary” context-dependence gives rise to some fascinating issues, which are taken up by Glanzberg in his (2001: section IV). Here are two (connected) issues. (1) If the domain of truth conditions is constrained by the resources speakers have for expressing propositions, then anti-realism about truth conditions might seem to follow. Glanzberg argues that it doesn’t. Just because certain truth conditions may be inaccessible from a given speaker’s context, they are in principle accessible to that speaker by a shift up the hierarchy—contexts are fully commensurable. And once a proposition is expressed, there is no reason to think that it does not satisfy bivalence, or cannot be evidence-transcendent, and so on. There is a sense in which truth conditions are constructed by speakers—but that need not force us to be anti-realist about truth conditions any more than we are forced to be anti-realist about buildings, or any other artifacts. (2) It is often thought that content is separable from context—once we settle on the content expressed, we can detach it from the context in which it was expressed. This is not so with extraordinary context-dependence, since there are propositions that just cannot be expressed in some contexts.

⁴³ Glanzberg draws on the theory of admissible sets (in Barwise 1975), and on Moschovakis’s work on acceptable structures (Moschovakis 1974).

⁴⁴ For further discussion of Glanzberg’s theory, see e.g. Gauker (2006).

29.8 SITUATIONS AND THE LIAR

For Glanzberg and Parsons, the key to a solution to the revenge liar is the context-dependence of quantifier domains—versions of *Repetition* and *Rehabilitation* are explained by a contextually-determined expansion of the background truth conditions. Barwise and Etchemendy also explain revenge in terms of a contextually-determined expansion—but of *situations*, not quantifier domains.⁴⁵

Barwise and Etchemendy employ two main tools, the notion of a situation taken from situation semantics,⁴⁶ and Aczel's set theory,⁴⁷ which provides a set-theoretical way of modeling circular propositions. A situation is a partial state the world might be in—for example the situation in which Claire has the ace of hearts. Now suppose I say on some occasion, "Claire has the ace of hearts." Following Austin, Barwise and Etchemendy claim that my statement expresses a proposition that involves two components: a *historical situation* (determined by "demonstrative conventions") and a *type* (determined by "descriptive conventions"). Thus all propositions include a contextually-determined feature, the situation they are about. And the proposition is true if the demonstrated situation is of the type described. Notice that if I am mistaken, and it's Dana at the table, not Claire, then the proposition is false, since the situation the proposition is about is not of the right type. And the proposition remains false even if Claire has the ace of hearts in a game across town.⁴⁸

Barwise and Etchemendy introduce *states of affairs* (SOAs) as the building blocks of situations. States of affairs are states of the world, and situations are represented set-theoretically as sets of states of affairs. The state of affairs where Claire has the ace of hearts is represented as $\langle H, c, a; 1 \rangle$, where H is the having relation, c is Claire, and a is the ace of hearts. If $\langle H, c, a; 1 \rangle$ belongs to a situation s , then Claire has the ace of hearts in the situation s . Conversely, if $\langle H, c, a; 0 \rangle$ —the *dual* of $\langle H, c, a; 1 \rangle$ —is in s , then Claire doesn't have the ace of hearts in s . The situation in which Claire has the ace of hearts and Max has the jack of clubs is the set of the two states of affairs $\langle H, c, a; 1 \rangle$ and $\langle H, m, j; 1 \rangle$.

The *type* of a state of affairs σ is represented by $[\sigma]$. So the type of the state of affairs in which Claire has the ace of hearts is represented by $[\langle H, c, a; 1 \rangle]$, which for convenience is written as $[H, c, a; 1]$. The *proposition* that Claire has the ace of clubs has two components: the situation s that is demonstrated (a particular portion of the world),

⁴⁵ Barwise and Etchemendy's central notion of a partial situation is drawn from the situation semantics of Barwise and Perry (1983).

⁴⁶ For situation semantics, see Barwise and Perry (1983).

⁴⁷ See Aczel (1987).

⁴⁸ On what Barwise and Etchemendy call a "Russellian" account of propositions, the proposition I express would be true. A Russellian proposition is uniquely determined by the sentence used, so if I say that Claire has the ace of hearts, and she does (albeit across town), then the Russellian proposition I have expressed is true. On the Russellian view, the situation I am talking about is not a feature of my utterance.

and the type of this situation. The proposition is represented by $\{s; [H, c, a; 1]\}$. And this proposition is true if s is of the type $[H, c, a; 1]$. Since we are modeling situations as sets of states of affairs, this proposition is true if and only if state of affairs $\langle H, c, a; 1 \rangle$ belongs to situations.⁴⁹

Now consider the liar. For each situation s , there is a liar proposition f_s which claims that its own falsity is a fact of s —where this fact is represented by $\langle Tr, f_s; 0 \rangle$. This “Austinian” liar is represented as $f_s = \{s; [Tr, f_s; 0]\}$. Barwise and Etchemendy go on to depict the role the actual world plays in the Austinian framework by introducing the notion of a *model of the world*. Once we have this notion and several other related notions on board, we can prove key results about the liar proposition f_s .

A *partial model* Ψ of the world is a collection of states of affairs satisfying the following three conditions:

- (i) No state of affairs and its dual are both in Ψ .
- (ii) If the state of affairs $\langle Tr, p; 1 \rangle$ is in Ψ , then the proposition p is true.
- (iii) If the state of affairs $\langle Tr, p; 0 \rangle$ is in Ψ , then the proposition p is false.

A situation s is *actual* in model Ψ if s is a subset of Ψ . A model Ψ is *total* if it is not properly contained in any other partial model.

We can now establish the following theorem:⁵⁰

Theorem Let s be an actual situation in some model. Then the liar proposition f_s about s is false.

Proof: If f_s is true, then $\langle Tr, f_s; 0 \rangle$ belongs to s . But s is actual in some model. So by the definition of a model of the world, f_s is false. Contradiction. So f_s is false.

In what follows, we fix a total model Ψ , and treat the notion of *actual situation* as relative to this model. And we need one further notion: a situation s is *F-closed with respect to a class of propositions* P if, for any false p in P , $\langle Tr, p; 0 \rangle$ is in s . We can now prove the following proposition:⁵¹

Proposition Let P be any set of propositions and let s be any actual situation which is F-closed with respect to P . Then the liar proposition f_s about s is not in P .

Proof: Recall that f_s is $\{s; [Tr, f_s; 0]\}$, so that f_s is true just in case $\langle Tr, f_s; 0 \rangle$ is in s . Suppose towards a contradiction that f_s is in P . Since s is F-closed with respect

⁴⁹ Barwise and Etchemendy assume some standard technique for representing distinct objects by distinct sets. Exactly how e.g. $\{\sigma\}$ and $\{s; [H, c, a; 1]\}$ are represented as sets will not matter, as long as they are represented differently in a systematic way.

⁵⁰ This is Theorem 6 in Barwise and Etchemendy (1987: 132).

⁵¹ This is Proposition 10 in Barwise and Etchemendy (1987: 135).

to P , then $\langle \text{Tr}, f_s; o \rangle$ is in s . So f_s is true. But by the above Theorem, f_s is false. Contradiction; so f_s is not in P .

Given this proposition, we can think of the liar sentence as providing a propositional function, from situations s to propositions f_s , that “diagonalizes out” of any set P of propositions. When the liar sentence “This proposition is not true” is used to make a statement about an actual situation s_1 , it expresses the false proposition f_{s_1} , but the fact that this proposition is false, represented by $\langle \text{Tr}, f_{s_1}; o \rangle$, *cannot be in the situation* s_1 (though it is in the fixed model Ψ). If we add this fact to s_1 , *then we get a different actual situation* s_2 . And the proposition $\{s_2; [\text{Tr}, f_{s_1}; o]\}$ is true, because $\langle \text{Tr}, f_{s_1}; o \rangle$ is in s_2 . This proposition says, about the expanded situation s_2 , that the liar proposition f_{s_1} is false—and this proposition is true. Similarly, the liar proposition f_{s_2} about s_2 is also false—but the fact that it’s false cannot be in the situation s_2 . But if we add this fact to s_2 to obtain s_3 , the proposition defined by $\{s_3; [\text{Tr}, f_{s_2}; o]\}$ is true. And so on, ascending a hierarchy of expanding situations.

In this way, a version of *Repetition* is explained. Given a liar sentence, say,

(λ) The proposition expressed by λ is false,

we can step back and recognize the falsity of (λ), consider that new fact, and say,

(μ) The proposition expressed by λ is false.

If we suppose that the original liar proposition is about a situation s_i , then the proposition expressed by the sentence λ is defined by $\{s_i; [\text{Tr}, f_{s_i}; o]\}$, and the fact $\langle \text{Tr}, f_{s_i}; o \rangle$ cannot be in s_i . But μ expresses a different proposition, since this proposition is about a different *extended* situation—given by $s_i \cup \{\langle \text{Tr}, f_{s_i}; o \rangle\}$ —where the fact that the proposition expressed by λ is false is added to s_i . And so the proposition expressed by μ is true. Both λ and μ are of the same type, but they are about different situations. This version of the revenge liar is resolved by a contextually-determined shift in what the two sentences are about—the domain of facts expands. Because of this expansion, μ can truly say something that λ cannot.⁵²

There are clearly close connections between Barwise and Etchemendy’s hierarchy of situations, Glanzberg’s expanding quantifier domains, and Burge’s hierarchy of indexed truth predicates. Glanzberg examines comparisons and contrasts between his account and Barwise and Etchemendy’s situation-theoretic account in Glanzberg (2004). Barwise and Etchemendy note that there are significant points of similarity between their theory and those of Parsons (1974) and Burge (1979). There are also parallels with

⁵² For further discussion of Barwise and Etchemendy’s theory, see e.g. Grim and Mar (1989); McGee (1991); Koons (1992); Glanzberg (2004); Gauker (2006).

Gaifman's pointer semantics (Gaifman 1988; 1992).⁵³ Similarities and differences between the accounts of Barwise and Etchemendy, Burge, and Gaifman are explored by Koons.⁵⁴

29.9 CHALLENGES

There are two major challenges to contextual theories of truth. The first is to show that the contextual approach is well-motivated. It is not obvious that the truth predicate is context-sensitive, or that the liar involves a contextually-determined shift in quantifier domains or situations, and so defenders of the contextual approach must meet the charge that the approach is ad hoc. The second major challenge is to address the threat of new paradoxes couched in terms of context—what about, for example, the sentence “This sentence is not true in any context”?

Perhaps enough has already been said about the first challenge. We saw that a contextual account of revenge fits easily into a familiar account of context and context-change, according to which we track context-change by keeping a running record of shifts in the salient information presumed to be available. In the course of the revenge reasoning, new salient information emerges—that the liar sentence is defective or pathological—and this produces a new context. Each of the contextual theories has somewhat different things to say about the effect on content of this contextual shift. According to both Burge's account and the singularity theory, the shift triggers the abandonment of the original truth schema and the adoption of a new one. For Glanzberg, the shift is an expansion of the salience structure which leads to an expansion of the domain of truth conditions. According to Barwise and Etchemendy, the new information allows us to step back and express a new proposition about a different extended situation. However these differences among the theories should be adjudicated, the contextual shift itself is well-motivated, drawing as it does on an entrenched account of context and context-change.

⁵³ Gaifman also starts out from the revenge liar, in the form: line 1: “The sentence on line 1 is not true.” line 2: “The sentence on line 1 is not true.” In the familiar way, we find the sentence at line 1 pathological, and so not true—and so we evaluate the sentence at line 2 as true. Gaifman draws this moral: we should assign truth values to tokens, not types, and the two tokens at lines 1 and 2 do not express the same thing. And that's because what these tokens express depends not only on their type, but also on the network of tokens with which each is associated. The associated networks for the two tokens are different, since the sentence at line 1 is self-referential and the sentence at line 2 is not: these tokens are *pointers* that point in different ways to the sentence at line 1. Pointers are the truth-bearers, and Gaifman develops a formal pointer semantics. Intuitively, the sentence at line 2 stands above the self-referential loop in which the sentence at line 1 is caught—so the evaluation of the pointer at line 2 involves a jump to a higher level. In general, pointers can be stratified in a way that reflects a Tarski-like hierarchy. Though Gaifman's theory is a theory about tokens, it isn't really a contextual theory of truth. The difference between the sentences at lines 1 and 2 is explained not in terms of a contextual shift, but rather in terms of the referential networks associated with each of the sentences.

⁵⁴ See Koons (1992: ch. 6).

Any account of the liar faces the second challenge: there is always the threat of new paradoxes for old. Perhaps the very terms of the solution give rise to new paradoxes. We've seen the challenge posed by the revenge liar: any theory of truth must say that the liar is defective in some way (it's ungrounded, or unstable, or indeterminate, or fails to express a proposition . . .)—but then a revenge liar that says of itself that it's defective in this way will be true. It is an advantage of contextual theories that they are not undermined by this form of revenge. The defectiveness of the liar sentence is not an absolute affair: assessment of the liar sentence breaks down in the initial context, but goes forward in the reflective context. In the case of R, for example, R has no truth_{CR} conditions, but is true_{TR}. And this accommodates revenge—indeed, as we noted at the outset, this form of revenge is a major motivation for contextual theories.

But perhaps there are other paradoxes that threaten contextual theories, paradoxes that are tailored to the essentials of each theory. Consider Burge's Tarskian account, according to which any occurrence of "true" is tied by context to a particular level of language. Then it may seem that a paradox is generated by the sentence "This sentence is not true at any level."⁵⁵ Burge responds that the attempt here to produce paradox is misguided—it tries, and inevitably fails, to "de-indexicalize" "true." Even in the phrase "true at some level," there is an implicit index on "true," so the attempt to quantify out the indexical character of "true" leads to incongruity (compare "here at some place").⁵⁶

A related challenge is posed by statements such as "All sentences are true or not"—how can such a global statement be accommodated if a use of "true" is always tied to a definite level? If we take such a statement to be asserted in a particular context, with a particular index on "true," then the broader import of the statement is compromised. In response, Burge distinguishes between *indexical* and *schematic* uses of "true." A predicate on an occasion of use is *indexical* if its extension depends on the context of use; it is *schematic* if it doesn't have a definite extension on that occasion, but through its use on that occasion provides general systematic constraints on the extension of the predicate on other occasions of use. Burge takes the formal principles of his theory to be stated schematically—and they are to be evaluated as true, where "true" is being used schematically.⁵⁷ Likewise, the global statement above is a schematic generalization. Its formalization is: $(s)(\text{Tr}_i(s) \vee \sim \text{Tr}_i(s))$, where the subscripts stand open, ready to be filled in as the occasions arise. And when we evaluate this schematic statement as true, we are using "true" schematically.

The singularity theory offers a very different response to these challenges, since the theory is not hierarchical and does not stratify the truth predicate. To fix ideas, let L be

⁵⁵ One might argue: either the sentence is true at some level, say *i*, or it is not true at any level. If the former, then, given what the sentence says, the sentence is not true at any level, including *i*—and we reach a contradiction. If the latter, then the sentence is true—and since truth is always tied to a level, it is true at some level. And we have a contradiction again.

⁵⁶ See Burge (1979: 108).

⁵⁷ Burge discusses schematic uses of "true" in Burge (1979) in Martin (1984: 115–17, 107–8), and in Burge (1982: 353).

a fragment of English that contains no semantic terms (and, further, for the sake of simplicity, no context-sensitive terms or vague terms). We obtain the language \mathcal{E} by adding to L the English predicate “true.” \mathcal{E} is the language that the singularity theory is a theory of. The key claim here is this: the language in which the singularity theory is couched—call it T —is *not* a Tarskian metalanguage for \mathcal{E} . A full defense of this claim would require a more detailed account of the singularity theory, but perhaps the broad outlines can be indicated here.⁵⁸

The main job of the singularity theory is to identify singularities of a given occurrence of “true,” and this is done via notions such as *determination set* and *primary tree*. The resources of the singularity theory are relatively meager. The theory requires only: (1) the notion of the type of an expression of \mathcal{E} ; (2) the notion of an utterance’s context; (3) the pragmatic information that a context is reflective with respect to a given expression; and (4) the determination set of a given expression. Nowhere does the language T of the theory contain a predicate coextensive with any occurrence of “true.” The theory does not provide a “model” of English. This is in contrast to say, a Tarskian account of truth which provides extensions of “true” at distinct levels, or Kripke’s theory, where the minimal fixed point is a model of the object language, providing the extension and anti-extension of “true.” There is nothing analogous in the language T of the singularity theory. T is a restricted theoretical language, free of context-sensitive terms, and since its sentences are not identified as singularities, every occurrence of “true” will have the sentences of T in its scope. Since we can regard T as a classical formal language, it is subject to Tarski’s theorem, and a Tarskian hierarchy can be generated from it. But since none of the theoretical sentences in this hierarchy are identified as singularities of “true,” the scope of the truth predicate of the object language \mathcal{E} arches over not only over T , but also over the languages of the hierarchy. In this strong sense, \mathcal{E} is expressively much richer than T . T is not a metalanguage for \mathcal{E} .

But neither is \mathcal{E} a metalanguage for T . From the context-independent perspective of T , it seems natural enough to quantify over contexts and form the predicate constant “sentence of \mathcal{E} true in some context.” In a sense, this predicate constant, call it “true _{\mathcal{E}} ,” can be understood as the truth predicate for \mathcal{E} —it applies exactly to those sentences of \mathcal{E} that are true in some context or other. (But, again, T is not a metalanguage for \mathcal{E} : any ordinary use of “true” is far more comprehensive in many ways than “true _{\mathcal{E}} .”) Now if we add “true _{\mathcal{E}} ” to T , then the following is a sentence of T : “This sentence is not true _{\mathcal{E}} .” This sentence says of itself that it’s not a true sentence of \mathcal{E} in any context. Since this isn’t a sentence of \mathcal{E} but a sentence of T , this is a true sentence of T . This true sentence of T won’t be identified as a singularity of any given occurrence of the context-sensitive predicate “true” of \mathcal{E} —so it will be in the extension of that occurrence. So ordinary uses of “true” will have in their extension sentences of T in which “true _{\mathcal{E}} ” occurs. There is no Tarskian hierarchy here. Yet since any given occurrence of the context-sensitive predicate “true” in \mathcal{E} has singularities, and these singularities will be true in some (reflective or neutral)

⁵⁸ Extended discussion can be found in Simmons (2010).

context, “true_ε” will apply to certain sentences that are beyond the scope of that particular occurrence of “true.” The extension of “true_ε” neither includes nor is included in the extension of any use of the context-sensitive predicate “true.”

From the point of view of the singularity theory, one can draw this lesson from the liar: there is an essential reciprocity between context-sensitive and context-independent perspectives. Each can do something that the other can't. On a given occasion of use, our ordinary context-sensitive predicate “true” comprehends *almost* all truths, including the truths of T and of any hierarchy generated from it—any use of “true” is as close to global as it can be. And though there are things that can be expressed in decontextualized theoretical languages that cannot be expressed via a given use of “true,” there is no one universal decontextualized language containing a global semantic predicate—one can always diagonalize out of such languages. We cannot take up both perspectives, context-dependent and context-independent, at once. But everything that can be said can be said from one or other of these perspectives.

Since Glanzberg's theory resolves the liar via an expansion of the quantifier domain, a *universal* domain of quantification cannot be admitted. Glanzberg argues for the general claim that there are no absolutely unrestricted quantifiers. It is an advantage of Glanzberg's approach that this response is independent of special considerations about the liar, and so escapes any charge of adhocness. Glanzberg takes it as truistic that meaning is a matter of interpretation, and that interpretation must provide a domain of quantification. The key question is whether it is possible for a speaker to specify a domain of “absolutely everything.” Usually domains are specified by using predicates, but what predicate could specify a universal domain? The predicate “object” might be suggested, but this predicate seems too vague to yield a determinate domain, and does not provide, by itself, a preferred sharpening (even a nominalist will not claim that it's part of the *meaning* of “object” that objects are concrete). The best hope for a maximally broad conception of object, Glanzberg suggests, is found in logic, via the logical notion of a singular term: an object is whatever a singular term refers to. Glanzberg gives his opponent this logical notion of object, but argues that it is still impossible to specify a domain of absolutely all objects—because of Russell's paradox. Given a specification of a domain, we can quantify over it and form the class term “{x:x=x}.” The class {x:x=x} cannot be in the domain over which x ranges, since if it were, the Russell set {x:x∉x} would be in the domain, by (restricted) comprehension. So we can never specify a domain of absolutely everything—something is always left out. And the argument need not be in terms of classes or sets: a general version of Russell's paradox can be formulated in terms of the notion of *interpretation*.⁵⁹ The process of interpretation itself can take us beyond the domain of any interpretation we produce. The logical notion of object is indefinitely extensible.

The claim that there are no absolutely unrestricted quantifiers is in line with most quantification in natural language. Quantifiers are usually restricted by predicates

⁵⁹ See Williamson (2003).

(“There’s no beer in the fridge”) or by the context (“Everything was destroyed in the fire”). Moreover, the extensibility of the logical notion of object is a special sort of expansion—the additional objects are artifacts of the process of interpretation, so that this expansion of the background domain has little practical effect on what we say. There are, however, cases where it might seem counterintuitive to give up on absolutely unrestricted quantification. Consider for example a logical truth such as “All objects are self-identical.” Glanzberg argues that this logical truth seems to be about absolutely everything because its truth does not depend on what the domain is. Though its quantifier must be interpreted as ranging over some domain, it tells us something beyond this: it tells us, in an ambiguous way, something about any domain it might be interpreted as ranging over. Such statements exhibit *typical ambiguity*: though its meaning is still fixed by its interpretation, with a specified domain, we can recognize that it would hold whatever domain was specified.⁶⁰ Logical validities provide one case of typical ambiguity. Global semantic statements provide another: any utterance of “Every proposition is true or not” comes with a contextually determined domain of truth conditions, but we can see that the statement would hold whatever the background domain. These special cases carry additional force—but that’s because they’re typically ambiguous, and not because they are genuine cases of unrestricted quantification.⁶¹

Just as Glanzberg argues that we cannot specify a domain of all objects, or all propositions, so Barwise and Etchemendy claim that no proposition can be about the world as a whole. Though (Austinian) propositions can be about extremely comprehensive situations, the falsity of a liar proposition, though a feature of the world, cannot be a feature of the situation the proposition is about. There is a hidden parameter that the Austinian account of the liar makes explicit: the part of the world that the proposition is about. Now in general the boundaries of the situation a person is referring to may well be unclear. Barwise and Etchemendy argue that this vagueness injects ambiguity into the everyday use of language—it is easy to think that the falsity of the liar proposition is part of the situation that the liar is about (contrast λ and μ above). This can make the liar seem intractable. But once we take proper account of these boundaries, we can draw the lesson of the liar: we cannot make statements about the whole world, about the universe of all facts.⁶²

Contextual theories respond to the threat of new paradoxes for old in a variety of ways, reflecting the variation among the contextual theories themselves. These responses

⁶⁰ Cf. Russell (1908), reprinted in Marsh (1971: 64–9). Relatedly, Parsons considers the following objection to his account in Parsons (1974): if we interpret the quantifiers of Parsons’s paper as ranging over some sufficiently large set, we can then produce a discourse to which his analysis of the liar will not apply. In response, Parsons suggests that the generality that his paper has, transcending any particular set as the range of the quantifiers, “must lie in a sort of systematic ambiguity, in that indefinitely many such sets will so” (Parsons 1974: 28 fn. 13). The notion of typical ambiguity is also related to Burge’s claim that there are schematic uses of “true.”

⁶¹ For critical discussion of the claim that there is no absolutely unrestricted quantification, see Williamson (2003) and the papers in Rayo and Uzquiano (2006).

⁶² See Barwise and Etchemendy (1987: ch. 13).

lead us to issues that go beyond the immediate scope of the liar. It may be claimed that this is an advantage of contextual approaches. There are no quick revenge liars that compromise contextual theories of truth. Instead we are led to an array of broad and related issues—issues about indexical and schematic uses of expressions, about the interplay between context-dependent and context-independent perspectives, about our ability to talk about absolutely everything. The hope is that by recognizing the context-dependence of truth, and taking on these broader issues, we can make progress toward a positive resolution of the liar.⁶³

REFERENCES

- Aczel, P. (1987). *Lectures on Nonwellfounded Sets*, CSLI Lecture Notes 14, Stanford, CA, Center for the Study of Language and Information.
- Allerton, D. J. (1978). "The Notion of 'Givenness,'" *Lingua* 44: 133–68.
- Antonelli, G. A. (1996). "Review of Keith Simmons, *Universality and the Liar*," *Notre Dame Journal of Formal Logic* 37: 152–9.
- Austin, J. L. (1950). "Truth," *Proceedings of the Aristotelian Society Supplementary Volumes* 24: 111–28. Reprinted in Austin, *Philosophical Papers*, ed. J. O. Urmson and G. J. Warnock (1961). Oxford: Oxford University Press, 117–33.
- Barwise, J. (1975). *Admissible Sets and Structures*. Berlin: Springer.
- Barwise, J. and J. Perry (1983). *Situations and Attitudes*. Cambridge, MA: MIT Press.
- Barwise, J. and J. Etchemendy (1987). *The Liar*. Oxford: Oxford University Press.
- Beall, Jc. (2003). "On the Singularity Theory of Denotation," in Beall (ed.), *Liars and Heaps*. Oxford: Oxford University Press, 253–61.
- Beall, Jc. and B. Armour-Garb (eds.) (2005). *Deflationism and Paradox*. Oxford: Oxford University Press.
- Brown, G. and G. Yule (1983). *Discourse Analysis*. Cambridge: Cambridge University Press.
- Burge, T. (1979). "Semantical Paradox," *Journal of Philosophy* 76: 69–98; reprinted with a post-script in Martin (ed.) (1984), 83–117.
- Burge, T. (1982). "The Liar Paradox: Tangles and Chains," *Philosophical Studies* 41: 353–66.
- Büring, D. (1999). "Topic," in P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press, 142–65.
- Chafe, W. (1976). "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View," in C. Li (ed.), *Subject and Topic*. New York: Academic Press, 25–55.
- Clark, H. and E. V. Clark (1977). *Psychology and Language*. New York: Harcourt, Brace, Jovanovich.
- Clark H. and S. Haviland (1977). "Comprehension and the Given–New Contract," in R. Freedle (ed.), *Discourse Production and Comprehension*. Hillsdale: Lawrence Erlbaum Associates, 1–40.
- Cresswell, M. (1972). "The World is Everything that is the Case," *Australasian Journal of Philosophy* 50: 1–13.
- Field, H. (2008). *Saving Truth from Paradox*. Oxford: Oxford University Press.

⁶³ My thanks to Michael Glanzberg for helpful comments on an earlier version of this chapter.

- Fintel, K. von (1994). *Restrictions on Quantifier Domains*. Ph.D. dissertation, University of Massachusetts at Amherst.
- Gaifman, H. (1992). "Pointers to Truth," *Journal of Philosophy* 89: 223–61.
- Gaifman, H. (1988). "Operational Pointer Semantics: Solution to Self-referential Puzzles I," in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, ed. M. Vardi. Los Altos, CA: Morgan Kaufmann, 43–60.
- Gauker, C. (2006). "Against Stepping Back: A Critique of Contextualist Approaches to the Semantic Paradoxes," *Journal of Philosophical Logic* 35: 393–422.
- Glanzberg, M. (2001). "The Liar in Context," *Philosophical Studies* 103: 217–51.
- Glanzberg, M. (2002). "Context and Discourse," *Mind and Language* 17: 333–75.
- Glanzberg, M. (2004). "A Contextual-Hierarchical Approach to Truth and the Liar Paradox," *Journal of Philosophical Logic* 33: 27–88.
- Gödel, K. (1944). "Russell's Mathematical Logic," in Schilpp (1944), 123–53.
- Grice, H. P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Grim, P. and G. Mar (1989). "On Situations and The World: A Problem for Barwise and Etchemendy," *Analysis* 49: 143–8.
- Grim, P. (1995). "Review of Keith Simmons, *Universality and the Liar*," *The Philosophical Review* 104: 467–9.
- Groenendijk, J., T. Janssen, and M. Stokhof (eds.) (1984). *Truth, Interpretation, and Information*. Dordrecht: Foris.
- Grosz, B. J. (1977). "The Representation and Use of Focus in a System for Understanding Dialogs," *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Los Altos, CA: Morgan Kaufmann, 67–76. Reprinted in B. J. Grosz, K. Sparck Jones, and B. L. Webber (eds.) (1986), *Readings in Natural Language Processing*. Los Altos, CA: Morgan Kaufmann, 353–62.
- Grosz, B. and C. Sidner (1986). "Attention, Intention, and the Structure of Discourse," *Computational Linguistics* 12: 175–204.
- Gupta, A. (1982). "Truth and Paradox," *Journal of Philosophical Logic* 11: 1–60; reprinted in Martin (ed.) (1984), 175–235.
- Halliday, M. A. K. (1967). "Notes on Transitivity and Theme in English: Part 2," *Journal of Linguistics* 3: 199–244.
- Halliday, M. A. K and R. Hasan (1976). *Cohesion in English*. London: Longman.
- Hardy, J. (1997). "Three Problems for the Singularity Theory of Truth," *Journal of Philosophical Logic* 26: 501–20.
- Heim, I. (1988). *The Semantics of Definite and Indefinite Noun Phrases*. New York: Garland.
- Isard, S. (1975). "Changing the Context," in Keenan (1975), 287–96.
- Juhl, C. F. (1997). "A Context-Sensitive Liar," *Analysis* 57: 202–4.
- Kamp, H. (1984). "A Theory of Truth and Semantic Representation," in J. Groenendijk, T. Janssen, and M. Stokhof (eds.) (1984), 1–41.
- Kamp, H. and U. Reyle (1993). *From Discourse to Logic*. Dordrecht: Kluwer.
- Kamp, H. and J. van Eijck (1996). "Representing Discourse in Context," in Johan van Benthem and Alice ter Meulen (eds.), *Handbook of Logic and Language*. Cambridge, MA: MIT Press/North-Holland, 179–237.
- Kaplan, D. (1989). *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein. New York: Oxford University Press.
- Karttunen, L. (1976) "Discourse Referents," in J. McCawley (ed.) *Syntax and Semantics* 7. New York: Academic Press, 363–85.

- Keenan, E. (ed.) (1975). *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press.
- Kleene, S. C. (1952) *Introduction to Metamathematics*. New York: Van Nostrand.
- Koons, R. C. (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge: Cambridge University Press.
- Kripke, S. (1975). "Outline of a Theory of Truth," *Journal of Philosophy* 72: 690–716; reprinted in R. L. Martin (ed.) (1984), 53–81.
- Kuppevelt, J. van (1995). "Discourse Structure, Topicality and Questioning," *Journal of Linguistics* 31: 109–47.
- Lewis, D. (1970). "General Semantics," *Synthese* 22: 18–67.
- Lewis, D. (1979). "Scorekeeping in a Language Game," *Journal of Philosophical Logic* 8: 339–59; reprinted in Lewis (1983), 233–49.
- Lewis, D. (1980). "Index, Context, and Content," in S. Kanger and S. Ohman (eds.), *Philosophy and Grammar*. Dordrecht: Reidel, 79–100; reprinted in Lewis (1998), 21–44.
- Lewis, D. (1983). *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.
- David L. (1998). *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.
- Martin, R. L. (ed.) (1984). *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press.
- McGee, V. (1991). "Review of Jon Barwise and John Etchemendy, *The Liar: An Essay on Truth and Circularity*," *The Philosophical Review* 100: 472–4.
- Muskens, R., J. van Benthem, and A. Visser (1997). "Dynamics" in J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*. Cambridge, MA: MIT Press/North-Holland, 587–648.
- Parsons, C. (1974). "The Liar Paradox," *Journal of Philosophical Logic* 3: 381–412; reprinted with a postscript in Martin (ed.) (1984), 9–45.
- Portner, P. and K. Yabushita (1998). "The Semantics and Pragmatics of Topic Phrases," *Linguistics and Philosophy* 21: 117–57.
- Priest, G. (2005). "Spiking the Field Artillery," in Beall and Armour-Garb (eds.) (2005), 41–52.
- Prince, E. (1981). "Toward a Taxonomy of Given-New Information," in P. Cole (ed.) *Radical Pragmatics*. Cambridge, MA: Academic Press, 223–55.
- Rayo, A. and G. Uzquiano (eds.) (2006). *Absolute Generality*. Oxford: Oxford University Press.
- Reinhart, T. (1981). "Pragmatics and Linguistics: An Analysis of Sentence Topics," *Philosophica* 27: 53–94.
- Roberts, C. (1996). "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics," *Ohio State University Working Papers in Linguistics* 49: 91–136.
- Russell, B. (1908). "Mathematical Logic as Based on the Theory of Types," reprinted in Russell, *Logic and Knowledge: Essays 1901–1950*, ed. R. C. Marsh (1971). New York: Putnam, 59–102.
- Scharp, K. (2007). "Replacing Truth," *Inquiry* 50: 606–21.
- Scharp, K. (2008). "Aletheic Vengeance," in Beall (2008), 272–319.
- Schilpp, P. A. (ed.) (1944). *The Philosophy of Bertrand Russell*. La Salle: Open Court.
- Simmons, K. (1993). *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.
- Simmons, K. (1994). "Paradoxes of Denotation," *Philosophical Studies* 76: 71–106.
- Simmons, K. (2000). "Sets, Classes and Extensions: A Singularity Approach to Russell's Paradox," *Philosophical Studies* 100: 109–49.
- Simmons, K. (2018). *Semantic Singularities. Paradoxes of Reference, Predication, and Truth*. Oxford: Oxford University Press.

- Stalnaker, R. (1975). "Indicative Conditionals," *Philosophia* 5: 269–86; reprinted in Stalnaker (1999), 63–77.
- Stalnaker, R. (1978). "Assertion," in P. Cole (ed.), *Syntax and Semantics*, vol. 9: *Pragmatics*. New York: Academic Press, 315–32; reprinted in Stalnaker (1999), 78–95.
- Stalnaker, R. (1988). "On the Representation of Context," *Journal of Logic, Language, and Information* 7: 3–19.
- Stalnaker, R. (1999). *Context and Content: Essays on Intentionality in Speech and in Thought*. Oxford: Oxford University Press.
- Stanley, J. (2000). "Context and Logical Form," *Linguistics and Philosophy* 23: 391–434.
- Vallduvi, E. (1992). *The Informational Component*. New York: Garland.
- Williamson, T. (2003). "Everything," *Philosophical Perspectives* 17: 415–65.
- Yablo, S. (2003). "New Grounds for Naïve Truth Theory," in Beall (2003), 312–30.

INDEX

- Absolute 127, 130, 131, 133, 134, 136, 137, 139, 143, 223
 absolute truth 135, 136, 143
 abstraction 111–12, 125, 140
 categorical 117
 in science 619–20
 acceptances 221, 225–6, 227, 231, 233–4, 235, 236, 692
 observational 232, 233
 actual world 272
 adequate ideas 81, 82–3, 223–4
 adequation theory 54, 56, 60, 62
 adequation to an object 45
 adverbialism 113
 adverbial truth 13
 aesthetic values 433, 434
 affirmations 34–5, 36, 37, 41, 42, 44, 45–6, 52, 175, 176, 203, 404
 affirmative existential judgments 107, 108, 109–10
 agent intellect 64, 67, 68
alēthēs/alētheia 9, 10–11, 12, 13, 14, 16, 20
 alethic being 11
 alethic disjunctivism 559, 560, 563–5
 algebraic accounts of propositions 316–17, 330
 analytic judgments of sense 134, 135, 136
 ancestral logic 43
 Andrić, M. 401–2
 Anselm of Canterbury 52
 anti-realism 219, 230–3, 357, 373, 436, 442, 443, 445, 461, 468, 503, 505 n. 5
 and science 614–15, 616–17
 and truth pluralism 544–6
 see also semantic anti-realism
 appearances 15–16, 17 n. 16, 18, 21–2, 233, 234, 235
 of objects 97–101, 105, 106
 approximate truth 614, 620–1, 622, 635 n. 18
 see also truthlikeness
 Aquinas, T. 45, 53, 54, 60, 94, 95, 96, 239
 cognition by intelligible species 62–4, 66, 68
 De Veritate 239
 Summa Theologiae 52, 239
 Aristotle 2, 4, 10, 14, 50, 51, 53, 54, 58, 60, 61, 64, 65, 69, 239, 396, 675, 677
 definition of truth and falsehood 36–7, 394, 395
 temporalist understanding of truth values 37–40, 41
 theory of the declarative sentence 33–5
 truth-bearers and truth-makers 39–44
 Categories 38, 40, 41 n. 85, 53
 De Anima 34, 35, 64, 239
 De interpretatione 33, 37 n. 74, 38, 39, 40 n. 83, 42 n. 85, 53, 239, 403
 Metaphysics 36, 37, 38, 40–1, 42–4, 45, 394, 403
 Nicomachean Ethics 33, 35, 61 n. 42
 Posterior Analytics 56–7, 61
 Prior Analytics 103
 arithmetic 656–7, 662–3, 710–11, 722
 Peano 413, 656, 663, 711, 720, 723–5, 727, 728, 729, 736
 Armstrong, D. M. 254, 267, 339, 340, 348, 349, 356, 359, 386, 642 n. 30
 assertions 40, 93, 103, 286, 287, 311, 312, 328, 691–2
 affirmative and negative 34–5, 36, 37, 42–3, 44, 45–6
 Frege on 209, 210, 211, 212
 relativist perspective on 534, 535
 superassertibility 455–6, 471, 472, 545–6, 556, 557, 570
 as truth-bearers 240
 warranted 230–1, 232–3, 294, 453–6, 470, 471, 546

- assertion-truth platitude 454
 assessment sensitivity 523–5
 atomic facts 163, 171, 243–4, 253, 379
 atomic propositions 163, 164, 165, 167 n. 12,
 171, 173, 571, 633, 635, 720
 atomic sentences 253, 254, 309, 310, 356, 440,
 441, 720
 atomism 17–18, 142
 logical 135, 253–5, 356, 378
 attitude(s) 507, 508
 de se 521, 525
 propositional 307, 322, 325–6, 692
 attributive truth 13, 20–1, 22, 23
 Augustine 50–1, 52, 53, 60, 62, 64, 65, 68, 675
 Austin, J. L. 1–2, 244, 245
 Avicenna 52
 axiom of choice 408
 axiom of consistency 733
 axiom of internal relations 127, 130, 131–2, 138,
 144, 145
 axiomatic theories of truth 6, 718–38
 base theory 719–20, 736
 formal 719–23
 typed truth 723–5
 see also type-free truth
 Ayer, A. J. 391 n. 32, 580–1, 582, 584, 587, 588,
 590, 688

 Baldwin, J. M. 422
 Baldwin, T. 272
 Barnard, B. 445 n. 23
 Barnes, E. 642 n. 30
 Barwise, J. 775–7, 778, 782
 Bealer, G. 316–17, 330
 Beall, J. C. 547 n. 7, 570, 752
 becoming 23 n. 37, 126
 Beebee, H. 355
 behaviorism 461
 being 13–14, 18, 52, 61, 65, 76, 77, 107, 111, 112,
 118, 126
 degrees of 23
 veridical 11–12, 17 n. 16, 36, 40, 41, 42, 43,
 44, 46
 see also not/non-being
 belief revision 645
 beliefs 24, 39, 40, 212, 221, 241, 270, 325, 488,
 489–90, 491–2

 affirmative and negative 42, 44
 doxastic 25, 26
 false 152, 153, 154
 incorrigible 227, 228
 Peirce on 285–7
 Ramsey on 185, 186
 relativist perspective on 535
 true 152, 153, 154, 185, 287, 291, 292, 293, 295,
 301, 307
 as truth-bearers 240
 see also judgment(s)
 Belnap, N. 722
 Benacerraf, P. 410 n. 10, 414, 655, 660 n. 18
 Berkeley, G. 434, 463, 464–5
 bias, in scientific work 609–10
 Bigelow, J. 349, 355, 356
 bivalence 38–9, 288, 357, 373, 394–8, 401, 403,
 443, 677
 and continuum hypothesis 407–10,
 414–19
 failures in set theory 410–14
 and future contingents 403–7
 and moral discourse 544, 546
 and realism 439–40, 441
 simple argument for 394–403
 vague statements 419–22, 430
 Blackburn, S. 450, 552 n. 14, 584
 Blanshard, B. 220 n. 1, 277, 278 n. 53
 body 82–3
 and mind 81
 Boethius, A. 53–4, 249 n. 15
 Bolzano, B. 94, 101–6, 109, 111 n. 23, 114,
 115, 117
 Bonaventure of Bagnoregio (St. Bonaventure)
 60–2
 Boole, G. 126, 201–2
 Boolean logic 201
 Bosanquet, B. 126, 132–3
 Bradley, F. H. 3, 125, 129–36, 137, 142, 143, 144,
 221, 239, 277, 278 n. 53
 on judgment 130, 132, 133–6
 regression argument 129–30, 132, 144
 on unreality of relations 126, 127, 129, 130,
 132, 133
 Appearance and Reality 126
 Essays on Truth and Reality 125
 The Principles of Logic 140

- Bradwardine, T. 677
- Brentano, F. 94, 101, 106–14, 115, 116, 117, 118
Psychology from an Empirical Standpoint
(PES) 107, 108, 109, 110, 115
The True and the Evident (TE) 109, 110, 111,
112, 113
- Broad, C. D. 244
- Brouwer, L. E. J. 408, 410, 466, 467
- Brown, L. 32 n. 63
- brute facts 79, 351
- brute truths 337, 338, 339, 345, 350–2
- bundle theory of objects 141
- Burali-Forti Paradox 414
- Burge, T. 208, 761 nn. 13, 14, 765–7, 777, 778,
779
- Burger, I. C. 633, 636
- Burgess, J. P. 505, 654
- Buridan, J. 59, 249 n. 14, 678, 680, 681, 691
- Burley, W. 54–6
- Burnyeat, M. F. 23 n. 39, 25 n. 43, 27 nn. 47, 48
- Byrne, D. 452
- Candlish, S. 260 n. 2, 272 n. 40, 277 n. 52
- Cantini, A. 733, 734 n. 18, 736
- Cantor, G. 407, 663
- Cappelen, H. 524, 531, 535, 536, 537
- capture 742, 745, 750, 752
- Carnap, R. 261, 473, 474, 650 n. 2
- Cartwright, R. 260 n. 2, 265
- Carus, P. 293
- categorical abstraction 117
- categorical acts 117
- categorical intuition 117, 118
- categoricity 661–3
- centered worlds 525–6
- certainty 57, 58–9
- Chihara, C. 512
- Chisholm, R. K. 272
- choice, axiom of 408
- Church, A. 391
- circumstances 521–2
- classical semantics 440
- clear and distinct ideas 78–9, 80, 82, 83
- cognition 54, 60–9, 152
by illumination 51, 60–2, 64, 65, 66, 67, 68
by intelligible species 62–4, 66–8, 69
and its objects 95–8, 102
Kant's theory of 95–8, 102
sensory 62, 64
- cognitive command 457, 458–60
cognitive shortcoming 458
- cognitive significance 391
- cognitive value 625, 645
- cognitivism 582–3
- Cohen, P. J. 408
- coherence theory of reality 221, 222, 223, 224,
229
- coherence theory of truth 3, 126, 128–9, 134,
136–9, 219–37, 278, 544
science and 604, 606, 608, 614–15
- colors 464
- combination and separation 34, 37, 40–1, 42,
43–4, 45–6
- commitment tables 593–7
- commitment theory 591–7, 598–9
- common-sense realism 463–4
- complete concept theory 87–8, 89–90
- complexes 151, 164, 167, 168–70, 172, 174, 178,
179, 180, 189
- complex senses 309
- compositional truth (CT) 724–5, 728
- compound propositions 164
mixed 549, 550, 561, 562–5, 568
- conceivability 469–70
- concept containment 83–4, 85–6, 87, 88,
90, 91
- concept expressions 308–9
- concepts 55, 62, 63, 68, 75, 140, 141, 142, 247,
260, 311
calculus of 201
as extensional 199, 200, 204, 207
as functions from objects to truth-values
195, 197–200
as intensional 199–200
as meanings 141
- conditionals 126, 196, 200, 202–3, 240, 298
material 747, 748, 749–50, 753
subjunctive 465, 466
- conformity 77, 98, 238, 242
correspondence as 239
- conjunctions 375, 425, 486, 488, 489, 490–1,
492, 493, 562, 564, 720
- consciousness 128
- consequence relations 488, 492–3

- consistency, axiom of 733
 constitution thesis 462, 463, 465
 constructive empiricism 613
 content(s) 114, 115, 117, 118, 197–9
 and context 757–9, 760, 764
 and fictionalism 507, 508, 509
 judgeable 198, 199, 202, 203, 204
 of judgment 105, 106, 107, 108, 109, 181, 189, 198
 pragmatist theory of 176, 177, 183–9, 190
 literal vs. figurative 512
 mathematical statements 657
 relativism 521–2, 524
 self-locating 525–6
 context 521–2
 of assessment 523–5
 and content 757–9, 760, 764
 and quantifier domains 771–4
 reflective status of 763
 of use 523, 524
 context-change 756, 757–9, 760, 778
 context-dependence 135, 136, 522, 537, 756, 774, 783
 contextualism 519, 520, 526, 529
 and disagreement 529–32, 538–9
 and eavesdropping 531–2, 539
 and epistemic modals 530, 531–2, 538, 539 n. 22
 non-indexical 524
 and predicates of taste 530, 531, 539
 and semantic paradox 755–86
 context of utterance 307–8, 318, 522
 contingent facts 58
 contingent liars 683–4, 685
 contingent truths 38–9, 42, 51, 88, 89, 335
 continuum hypothesis 407–10, 414–19, 663
 contraction 743, 744, 745
 contradiction 35, 677–9, 721, 722, 723, 724, 729
 Convention T 697, 698, 699 n. 14, 700–1, 704, 705–6, 715, 720, 721, 722
 convergence, scientific knowledge 610–12
 conversational score 758–9
 Cook, R. 683
 copula 34, 36, 56, 104, 105, 107, 117, 130, 193
 core logic 357
 correctness 94, 112, 113, 114, 118, 230
 in mathematics 648, 649–50
 correlation 245–6
 correspondence 77, 103, 226, 227, 261–2, 457, 478, 496, 498, 501, 514, 515, 545, 546, 556, 557, 570
 as conformity 239
 as congruity or isomorphism 46, 245, 246
 as correlation 245–6
 correspondence theory of truth 3, 36, 126, 134, 166, 219–20, 233, 235, 238–58, 348 n. 26, 434–5, 454, 455, 544
 Aristotle and 45–6
 fact-based versions 239, 241–4, 248, 249
 Joachim and 136–7
 Moore and 142, 175, 238, 239, 243, 245
 non-identity principle 245, 251
 object-based versions 239–40, 248, 249, 255
 pragmatism and 283, 284, 285
 Ramsey and 182
 Russell 147–8, 150, 151, 155, 159–60, 161–2, 163, 168, 172, 175, 182, 189, 190, 238, 239, 244, 251–3
 science and 604, 606, 607, 609–10, 611, 616, 618, 621
 structure and 246, 248, 249–53
 Tarski and 696
 Wittgenstein 151, 162–5, 167, 172, 190, 243
 cosmological role of facts 457–8, 460
 co-supposition 71–2
 Cotnoir, A. 547 n. 7, 555 n. 22, 571 n. 55
 counterpart theory 342–4
 creative extension 391
 credal states 644
 Cresswell, M. 763
 Crivelli, P. 29 n. 54, 35, 37 n. 74, 42, 43 nn. 90, 91
 Currie, G. 511
 Curry's paradox 677, 727, 748, 749
 cut 743, 744, 749–50, 751

 David, M. 262 n. 10, 265 n. 17, 267 n. 23
 Davidson, D. 219, 225, 255, 708 n. 40, 723
 decidable statements 439, 450
 declarative sentences 33–5
 deflationism 4, 133, 184, 212–13, 268, 273, 299, 427, 477–502, 504, 510, 515, 582–4, 599, 621, 719, 725
 deflationist theory of truths 493–7, 500, 501

- double-counting objection to truth
 pluralism 551–2
 semantic-descent 481, 483–4, 485, 486, 487,
 488, 494, 498, 500
 truth-ascription 488–93, 496, 497, 498, 500
 T-schema 480, 481–3, 484–5, 487, 488, 494
 Wrights's critique of 452–6
- degrees of truth 143, 626
- Democritus 17–18, 674
- demonstratives 318
- De Morgan, A. 126
- denials *see* negation
- deontic modals 527
- Descartes, R. 57, 59 n. 36, 75, 76–80, 82, 83–4,
 86, 88, 90, 95, 222, 223, 227
- descriptive semantics 448
- de se* attitudes 521, 525
- determinacy 4, 418, 444, 460, 544
 of sense 398–403
- determination pluralism 557–8, 560, 563–5
- deviant logic 687–8
- DeVidi, D. 713
- Devitt, M. 438, 462, 463, 464, 465, 466, 468,
 503
- Dewey, J. 283, 284, 293–7, 299, 300, 302
- Dickson, G. 671
- difference 31, 32
- dilemma, rule of 357
- direction terms 441–2, 460
- direct reference 318–19
- disagreement 592
 contextualist explanations of 529–32,
 538–9
 faultless 532–4, 535–6, 539
 relativist explanations of 529–34, 535–8
- disjunctions 38, 167, 240, 356, 423–4, 425,
 426–7, 429, 430, 564
- disjunctive facts 163, 166, 167 n. 12
- disjunctive propositions 164
- disjunctivism 262 n. 11, 273–4, 279
 alethic 559, 560
- Disquotational Schema 554
- disquotational theory 721, 726–8
- distance 634–41, 643
- divine
 illumination *see* illumination
see also God
- Dodd, J. 264 n. 15, 265 n. 17, 268, 269, 274, 275,
 355
- domain (of a function) 361, 362, 363, 366, 367,
 368, 369, 370, 377, 379, 382, 383, 385,
 388, 391, 392
- domain (of truth conditions) 773–4, 778
- doubt 285–6, 293, 307
- Dowell, J. 531
- doxastic beliefs 25, 26
- Dreier, J. 537, 585–6, 587, 588
- Ducasse, C. J. 261
- Duhem, P. 617
- Dummett, M. 4, 176, 195, 196, 204, 266 n. 21,
 355, 373, 408, 422, 435–6, 437, 460,
 461–2, 466–8, 469, 473, 503, 523 n. 9,
 655, 661
 on bivalence 405–7
 and coherence theory of truth 219, 230–3
 manifestation argument for semantic
 anti-realism 446–52
 on phenomenalism 465–6
 semantic criteria for realism 438–46, 460
Thought and Reality 232
Truth and the Past 232
- Duns Scotus, J. 60, 66–8
- early analytic philosophy *see* Frege; Moore;
 Ramsey; Russell; Wittgenstein
- eavesdropping 531–2, 537–8, 539
- Edwards, D. 546 n. 6, 559 nn. 34, 35, 571 n. 55
- Egan, A. 519, 521, 525–6, 527, 531 n. 16
- einai* (to be) 11, 11–12 nn. 5, 6, 13–14, 30
- Eklund, M. 512, 513, 547 n. 6
- Eleatics 22
- elementary propositions 184, 185, 187
- Elgin, C. 692
- emotivism 580–1, 588
- empirical-object discourse 498–9
- empiricism
 constructive 613
 feminist 609
 logical 614
 and science/scientific knowledge 603, 609,
 613, 614
- endorsement view 485–6, 491, 499
- entailment principle 375, 509
- entity realism 615–16

- Epimenides 674–5
epistēmē 19–20, 26, 56
 see also science/scientific knowledge
epistemicism 444
epistemic modals 526, 527, 528, 530, 531–2,
 538, 539 n. 22
epistemic truth 12–13, 16, 20
epistemic utility theory 644, 645
epistemology 2
Ericsson, K. A. 674
error theory 434, 436, 438, 462, 470, 471, 472–3,
 503, 504, 505, 507, 509, 511 n. 16
essentialism 334
essential richness thesis 708, 713–14
Etchemendy, J. 698 n. 10, 775–7, 778, 782
eteos/etymos/etētymos 10
ethics 5, 231, 236, 295, 296, 300
Eubulides 676
Euler, L. 650–1
Euthyphronic discourse 457, 460–1 n. 25
evaluation rules, truth-maker/falsity-maker
 365, 366–71
events 158, 180
evidence 113, 114, 118–19, 230, 231, 232, 233,
 236, 286, 294
 indirect 233
 potential 232–3
evident judgments 112, 113–14, 118
evidentness 57, 58–9
excluded middle, law of 36, 38, 39, 220, 357,
 403, 404, 408, 410, 430, 443, 459, 466,
 467, 687, 729, 746, 749
exclusionary semantics 184, 187–8
exemplars 65–6, 67
exemplification 692–3
existential facts 254
existential generalization 316
existential judgments 107–8, 109–13
existential propositions 141–2
experience 287, 288, 294
 ideal 289
 and ideas 79, 140
 possible 233
 real 289
 of truth 106–14, 118–19
expertise 674
explanation 487–8, 493–7
explosion 747, 749
expressiveness 487, 496
expressivism 436, 438, 462, 470, 535 n. 20,
 579–80, 583, 586–8
 and commitment theory 591–7,
 598–9
 and Frege-Geach problem 588–92, 598
extension(s) 199, 200, 204, 207, 315, 316
external relations 131, 138, 139, 143–4, 152
facts 3, 45, 56, 101, 114, 125, 127, 133, 135, 137, 138,
 140, 145, 152, 153, 154, 158, 160, 165–76,
 180, 182, 219–20, 226, 238, 255–7, 260
atomic 163, 171, 243–4, 253, 379
brute 79
chunk conception of 165, 167, 168, 169, 171
conceivable 469–70
constituent structure 256
contingent 58
cosmological role of 457–8, 460
disjunctive 163, 166, 167 n. 12
existential 254
Fregean 266–9
as incomplete symbols 173–6, 176
moral 457
negative 163, 164, 166, 167, 173, 254
objective 94
physical 457
positive 173
temporal 12, 16
and true propositions, identity between
 260–79
and truth values 37–8
universal 254
Wittgenstein on 163, 164, 165–76
false beliefs 152, 153, 154
false circle 24 n. 42
falsehood(s) 10, 24, 25, 35, 43–4, 45, 76, 84,
 145, 241, 244
absolute worthlessness of 628, 631, 643
necessary 41
objective 153, 157
Plato's account of 26–32
relative worthlessness of 628, 631, 632, 643
and truthlikeness 625, 626, 627, 632, 645
value of content for (VCF) 628, 630–1, 639
false ideas 76, 79–80, 81–2, 84

- false judgments 155, 156, 157, 181, 224
false propositions 271–2, 273, 274, 275, 312,
313, 325
Bolzano on 103
Moore on 141, 142
Russell on 153, 155, 158–9, 160, 161, 178, 179,
189, 221, 252
falsification 627
and vague statements 424, 425, 426
falsity-makers 363, 364, 382, 383, 386–7, 389,
390, 391, 392
evaluation rules 365, 366–71
formal co-inductive definition of 379–81
Feferman, S. 733, 734, 735
see also Kripke-Feferman axiomatic theory
feminist critiques of science 604, 608–10, 618
feminist empiricism 609
Ferejohn, M. T. 32 n. 63
Feyerabend, P. 517, 605
fictionalism 447, 448, 449, 503–16
about truth(s) 512–15
hermeneutic 505, 506, 507, 508, 509, 510,
511, 514, 515
and mathematics 652, 657
modal 506, 515
paraphrase 506–7
possible worlds 506–7, 515
prefix 505–8, 509, 510–11, 512, 514, 515
pretense 507, 508, 509, 510, 511–12,
514, 515
revolutionary 505, 506, 507, 508–9, 510, 511,
512, 514, 515
fictions 4, 145, 153, 160, 471–2, 473, 498
historical 504
natural view of 504
realistic 504
Field, H. H. 436, 462, 471, 504–5, 513,
570 n. 53, 651–2, 715, 737, 756
figuralism 507, 512, 515
Fine, K. 265, 269 n. 32
finitism, meaning 608
Fintel, K. von 537–8, 539 n. 22
Fish, W. 273 n. 43, 275
Fitch, F. 261
formal distinction 66–7
formal truth 98
Forms, Plato's theory of 15, 18–24, 25, 51
foundational semantics 448
Fox, J. 355, 356
Fraassen, B. van 505, 612–13, 734 n. 18
Francén, R. 537
Franssen, M. 642–3
Frede, M. 31
Fregean facts 266–9
Fregean propositions (thoughts) 210–11,
212–13, 249, 250, 254, 257, 266–7, 268,
269, 308–11, 330
Frege, G. 3, 108 n. 19, 130, 140, 141, 153, 171,
181, 193–215, 245, 247, 378, 437, 440, 441,
442, 445, 469, 473
and bivalence 439
concepts as functions from objects to
truth-values 195, 197–200
and deflationism 212–13
identity of facts and true propositions
261–2
and mathematics 648, 650, 651, 652, 653,
654, 660–1
regress argument 208–12, 262 n. 9, 266
n. 22
sentences as referring to truth-values
195–7, 203–4
sentential connectives as truth-functions
200–4
truth-values as objects 204–8
Begriffsschrift 197, 200, 201, 202–3, 205, 211
Function and Concept 200
Grundgesetze der Arithmetik 132, 193, 194,
196, 204, 205, 206, 207, 208
“On Sense and Reference” 204, 209, 210
Frege-Geach problem 588–92, 598
Friedman, H. 730
functionalism
manifestation 557, 560, 562–3
second-order 556, 560, 562
future 232, 461
future contingents 38–9, 403–7
relativism 527–8
future tense statements 443–4, 445
Gaifman, H. 778
Game Analogy Argument, and truth
pluralism 546
game formalism 650

- Gaskin, R. 261, 272, 273 n. 41
 Geach, P. 203, 688
 see also Frege-Geach problem
 Gemes, K. 634
 generalization
 devices of 482, 483–4, 501
 existential 316, 679
 problem of 552 n. 15, 696, 706
 generic truth 560, 561, 567
 Gentzen, G. 376
 Gerla, G. 630, 632 n. 14, 636 n. 19, 639
 Gibbard, A. 586, 587–8, 592
 Giere, R. 621
 Gillies, A. 537–8, 539 n. 22
 Gilson, E. 62
 Glanzberg, M. 531, 535, 686, 771–4, 775, 777, 778, 781–2
 God 51, 52, 59, 62, 64, 79, 80–1, 89, 222, 223, 224, 233, 290
 belief in 290
 and Malebranchian occasionalism 582
 Gödel, K. 193–4, 407–8, 408, 415, 686, 720, 767
 incompleteness theorems 194, 655, 663, 685, 709
 Goldbach's Conjecture 408–9, 410, 413
 Goldblatt, R. I. 427
 Goldstein, L. 681
 Goodman, N. 692–3
 the Good 23, 51
 Gorgias 17
 Graham, A. C. 672
 Green, T. H. 127–9, 132, 137
 Grelling, K. 686, 693
 Grosz, B. J. 759, 760
 groundedness 755–6, 765–6 n. 25
 grounded truth 116, 334, 335, 563, 691, 733
 Gupta, A. 709 n. 46, 722, 756

 Haack, S. 688
 Hacking, I. 283
 Hägerström, A. 580 n. 1
 Halbach, V. 728, 730, 733, 736
 Hale, B. 441, 449 n. 14
 Hanks, P. 321, 326–9
 Hanson, N. R. 605
 Hare, R. M. 590

 Harman, G. 264, 267 n. 26, 269, 517–18, 540
 Hawthorne, J. 527, 531, 535, 536, 537
 Hegel, G. W. F. 101, 126
 Heidegger, M. 10, 117
 Heidema, J. 633, 636
 Heil, J. 340–1, 342, 349
 Heim, I. 759, 760
 Hein, P. 690
 Heitsch, E. 10
 Hellman, G. 656
 Henry of Ghent 60, 64–6, 67, 68
 Heraclitus 16
 Herbert, E. 126
 hermeneutic fictionalism 505, 506, 507, 508, 509, 510, 511, 514, 515
 Herzberger, H. 683, 687
 Hestir, B. 20 n. 24
 Heyting, A. 408
 hierarchy 764–7, 780
 Hilbert, D. 651
 Hilpinen, R. 635, 636
 Hintikka, J. 640
 historical fictions 504
 historicist approaches to scientific knowledge 604–6, 611–12
 Hobbes, T. 76, 79, 84–5, 87, 88, 91
 Hölderlin, F. 101
 holism 142, 278, 283, 288, 299–300
 meaning, in science 606
 Homer 10
 Iliad 10 nn. 3, 4
 Odyssey 10 nn. 3, 4
 Horgan, T. 555–6 n. 23
 Hornsby, J. 275, 276, 277, 279
 Horwich, P. 212, 699 n. 13, 727, 728
 Hume, D. 76, 84 n. 13, 127, 238
 Hume's Principle 660
 Husserl, E. 94, 114–20, 242, 251
 Huvenes, T. 539
 hypothetical judgments 134

 ideal entities 115–16
 ideal experience 289
 idealism 2, 3, 93–120, 219, 223–4, 279
 Berkeleyan 434, 463, 464–5
 British 125–49, 239, 277–8
 Kantian *see* Kant

- idealized models in science 619, 620
ideal language 248, 250
ideas 75, 78
adequate 81, 82–3, 223–4
agreements and disagreements between 85–6
Bradley on 133
clear and distinct 78–9, 80, 82, 83
complex 86
Descartes on 78–80, 82
and experience 79, 140
false 76, 79–80, 81–2, 84
inadequate 82
innate 79–80
joining and separating of 85
Locke on 77, 78, 79, 84, 85–7
logical 140
phenomenology of 79–80, 82
simple 86
Spinoza on 81–3
true 76, 77, 78, 79–80, 81, 82, 83, 84
identity theory of truth 3, 32, 85, 88,
109 n. 21, 142–3, 147, 148, 175, 176,
206, 251, 259–82
illumination 51, 60–2, 64, 65, 66, 67, 68
imagination 67, 289, 296
imperative propositions 327
inadequate ideas 82
incoherence view of truth 705, 706
incompleteness theorems (Gödel) 194, 655,
663, 685, 709
inconsistency theory 512–13
indefinability theorem 396, 397, 703, 704, 708,
709, 712–13
indeterminacy 25, 400, 422, 426, 444, 445,
459, 460
indexical judgments 134–5, 136
indexical relativism 524
indexicals 134–5, 311, 318, 319, 520, 521, 523,
526, 762
index semantics 314
induction
mathematical 664
pessimistic 618
inexpressibility theorem (Tarski's theorem)
708, 709–12
inference 87, 111, 126, 194, 227, 229, 231
mixed, and truth pluralism 550–1, 561, 562,
565, 568–9
scientific 617
inference-licensing property 592–3, 595
innate ideas 79–80
inquiry 285–9, 292, 293–4, 295–6, 302
instantiated propositions 325
instrumentalist view of science 614, 615 n. 7,
621, 622
intellect 60, 62, 64, 95, 239
agent 64, 67, 68
and reality 54, 56, 62
and will 83
intellectual acts 68, 69
intellectual habits 68, 69
intelligible species 55, 62–4, 66–8, 69
intensional model structures (IMS) 316
intension(s) 199–200, 315
intentionality 114, 247
intentional objects 77
intentions 63
internal relations 127, 130–2, 138, 143–4, 145,
359
interpretation 781
interrogative propositions 327
intuition 86, 97, 100, 111, 117
categorical 117, 118
intuitionism 407–10, 462, 466, 467
intuitionistic logic 4, 357, 421–2, 570
isomorphic correspondence 46, 245, 246
Jackson, F. 531 n. 16
Jackson, J. 472
James, W. 283, 284, 290–3, 294, 295, 298, 299,
300–1, 302, 611
Jevons, W. S. 126
Joachim, H. 126, 136–9, 148, 221, 223–4, 225,
233, 239, 277, 278 n. 53
John's Gospel 51
judgment(s) 24, 25–6, 27, 33, 39, 46, 65, 83, 86,
95, 99, 100–1, 102, 103, 126, 127, 139, 143,
221, 241, 328
acts of 107, 108, 115, 117
analytic judgments of sense 134, 135, 136
Bradley on 130, 132, 133–6, 140
Brentano on 106–14
calculus of 201

- judgment(s) (*cont.*)
 content of 105, 106, 107, 108, 109, 181, 189, 198
 pragmatist theory 176, 177, 183–9, 190
 evident 112, 113–14, 118
 existential 107–8, 109–13
 false 155, 156, 157, 181, 224
 Frege on 209–12
 hypothetical 134
 indexical 134–5, 136
 necessary 142
 and object, harmony between 110
 predicative 27, 32
 and representation 107, 108–9
 Russell on *see* Russell
 true 155, 156, 157, 159, 181, 239, 239–40
 as truth-bearers 240
 unity of 130, 132, 134
see also beliefs; multiple-relation theory of judgment
- Kahn, C. 11–12
- Kant, I. 75, 91, 94, 95–101, 102, 107, 108, 117, 118, 119, 127, 128, 222–3, 231, 233, 234–5, 473, 474, 652
Critique of Pure Reason 91, 95, 96, 97, 98, 99–100
Prolegomena 97, 98, 99
- Kaplan, D. 252, 311, 318, 319, 521, 522, 523 n. 10
- Keller, L. J. 329, 330
- Keynes, J. N. 126, 139
- King, J. C. 320–3, 329, 330
- Kirkham, R. L. 447, 696 n. 3, 708
- Kleene schema (strong and weak) 594, 595, 731, 732, 733–4, 748 n. 9, 769 n. 31
- knowing 12–13, 19, 55, 61
- knowledge 19–20, 61–2, 85, 86, 115, 152, 227, 234, 718
 imperfect 81–2
 scientific *see* science/scientific knowledge
- knowledge ascriptions 527
- knowledge of meaning
 fictionalist view of 447, 448, 449
 implicit 447
 substantial view of 446–7, 447–8, 449
- Kölbel, M. 519–20, 521 n. 8, 527, 533, 534
- Kolodny, N. 527
- Koons, R. C. 764, 778
- Kreisel, G. 414–19, 661
- Kripke-Feferman axiomatic theory 730, 732–4, 734 n. 19
- Kripke, S. 194, 311, 317–18, 360, 448, 513, 674, 709 n. 46, 722, 725, 733, 734 n. 18, 755–6, 765 n. 25, 766, 780
- Kuhn, T. 517, 605–6, 608, 611–12, 614
- Kuipers, T. 630–1
- Künne, W. 36 n. 71, 239
- language 6, 9, 39, 224, 225, 227, 228, 230–1
 ideal 248, 250
 and truthlikeness 641–2
see also metalanguage; natural language
- Lasersohn, P. 520, 521–2, 524, 527, 530
- Leibniz, G. W. 76, 87–90, 91, 95, 127, 688
- Leitgeb, H. 730
- Lesniewski, S. 696 n. 2, 697 n. 7, 698
- Lewis, C. I. 284, 299
- Lewis, D. 314, 315 n. 25, 343, 444, 510, 511, 521, 522–3 n. 9, 525, 758–9
- liar paradox 5–6, 154, 397, 512, 547, 599–600, 671, 674–5, 676–7, 678, 681, 685, 688–9, 726, 734, 739, 742–5
 infinite 682–3
 looped 680, 682
 Tarski and 686, 687, 696, 698, 702–4, 706
see also revenge liar
- Locke, J. 76, 77, 78, 79, 83–7, 88, 127–8
- logic 2, 4, 5–6, 54, 126, 194, 195, 231, 240, 288, 289, 300
 ancestral 413
 Boolean 201
 core 357
 deviant 687–8
 intuitionistic 4, 357, 421–2, 570
 non-classical 739–54
 noncontractive 750–2, 753
 nontransitive 749–50, 753
 quantum 427, 428–9
 three-value 688–9
 of vague statements 422–7
- logical atomism 135, 253–5, 356, 378
- logical consequence, and truth-makers 359, 360, 371–5, 383–6
- logical constants 185, 188, 253

- logical empiricism 614
 logical ideas 140
 logical pluralism 570
 logical positivism 604–5
 lottery paradox 645
 Lowe, E. J. 355, 364 n. 13, 375
 Löwenheim-Skolem Theorem 385 n. 29, 413
 Lucas, J. R. 671
 Łukasiewicz, J. 688
 Luther, W. 10
 Lynch, M. 544 n. 2, 545 n. 4, 546 n. 6, 547 n. 7,
 549 n. 10, 553, 555 n. 22, 556 n. 24, 556–7
 nn. 26–30, 563 n. 40, 569 nn. 50, 51

 McCarthy, D. 140
 MacColl, H. 201 n. 17
 Macdonald, C. 275
 McDowell, J. 259 n. 1, 262, 273 n. 43, 274–5,
 276, 277, 279, 452
 MacFarlane, J. 53, 520, 521, 523–5, 527–8,
 529–30, 532, 534, 536–7, 539–40
 McGee, V. 413–14, 727–8, 730
 McGinn, C. 449, 450
 Mackie, J. L. 262 n. 8, 434, 436, 462, 471
 Mackinson, D. C. 675
 Malebranchean occasionalism 582
 manifestation argument 446–52
 manifestation functionalism 557, 560, 562–3
 Mares, E. 752
 Martin, R. L. 764
 Marty, A. 108 n. 19, 111 n. 23
 material conditionals 747, 748, 749–50, 753
 material supposition 70
 mathematical objects 462, 466–8, 653–4, 657
 abstract 648, 649, 654, 658
 Fregean argument for existence of 648,
 653, 654
 and mathematical structures 661
 platonist view of 648–9, 658, 659–61
 mathematical statements 495–6, 497, 498,
 648
 content of 657
 truth-value realism 649, 658–64
 mathematics 5, 18, 160, 231, 236, 288–9, 300,
 435–6, 461, 648–65, 722
 correctness in 648, 649–50
 and fictionalism 652, 657

 formalist approach to 649–50
 Fregean approaches to 648, 650, 651, 652,
 653, 654, 660–1
 non-literal language 657–8
 objectivity in 648–9, 658–64
 proof of theorems 649–50, 655
 semantics for language of 653–8
 theoremhood 649–50, 655
 see also arithmetic; set theory
 Maudlin, T. 691–2
 meaning 509, 781
 finitism 608
 and pragmatism 611
 scientific knowledge 606, 608
 speaker 506, 510
 theory of 126
 see also knowledge of meaning; semantic;
 semantics
 medieval tradition 50–74
 Meier, G. 95
 Meinong, A. 110, 272, 273 n. 41, 442
 Meinongianism 243
 Melia, J. 346
 Melissus 16 n. 14
 Mellor, D. H. 184, 340, 341, 342
 mental states 433–4, 461, 464–5, 590, 591, 592
 Menzel, C. 316, 414
 Merricks, T. 329, 330
 metalanguage 357, 362, 702, 708, 709, 713–14,
 722, 725, 726, 729, 736, 780
 metaphors 462–3, 466, 468, 657, 658
 metaphysical status of truth 51–3
 Miller, A. 438, 452, 462, 463–4, 466, 468, 472
 n. 31
 Miller, D. 628, 629, 630, 639
 Miller-Tichý incomparability proof 628, 632
 Mill, J. S. 238, 249 n. 15
 mills, E. 678
 mind 75, 97
 and body 81
 mind-dependent/mind-independent
 existence 571
 minimalism 437, 452, 454–6, 459, 460, 470–1,
 546 n. 5
 misidentification 27
 misjudgment 27
 mixed atomics 550 n. 11

- mixed discourse, and truth pluralism 549–51,
 561–5, 568–9
 modal fictionalism 506, 515
 modality 313–14, 342, 343–5, 444, 469
 deontic 527
 epistemic 526, 527, 528, 530, 531–2, 538,
 539 n. 22
 modal realism 444
 modal structuralism 656–7, 659
 models, verification and falsification of
 sentences in 362–5
 moderate truth pluralism 553, 554 n. 18,
 555–60
 alethic disjunctivism 559, 560, 563–5
 determination pluralism 557–8, 560, 563–5
 and double-counting objection 566
 manifestation functionalism 557, 560,
 562–3
 and mixed discourse 561–5
 response to instability challenge 560–1
 response to unity challenge 561
 second-order functionalism 556, 560,
 562
modus ponens 194, 492, 590, 591, 598
 Moffett, M. 264 n. 14
 Molnar, G. 346, 347
 monads 87, 89
 monism 130, 278, 543, 553
 Montague, R. 314
 Moore, G. E. 2, 3, 109, 110, 113 n. 28, 125, 139–44,
 145, 148, 153, 249 n. 15, 251, 269, 674
 and correspondence theory of truth 142,
 175, 238, 239, 243, 245
 criticism of James 292–3
 moral discourse 456–7, 470–1, 498, 544, 545
 moral facts 457
 moral inquiry 296
 moral relativism 517–19
 moral truth 14, 20, 579–601
 moral values 433, 434
 Mormann, T. 630, 639
 Mortensen, C. 632 n. 14
 Mostowski, A. 722
 Mo-tzu 672–3
 Mourelatos, A. 25 n. 43
M-relative truth-makers 365–6
 Mulligan, K. 355, 356, 378
 multiple-relation theory of judgment
 Ramsey 177–80, 181
 Russell 127, 145, 146–7, 148, 151, 152, 154–62,
 176, 178, 189, 190
 Wittgenstein 172, 184, 190
 Murzi, J. 752
 Myhill, J. 709

 Nagel, E. 294
 naturalism 300
 natural kind terms 448
 natural language 212, 307, 308, 314–15, 322,
 481, 484, 486, 520, 526, 528, 540, 586,
 591, 657, 685, 686, 703, 704–5, 725, 742,
 765, 767
 Neale, S. 256
 necessary falsehoods 41
 necessary judgments 142
 necessary truths 41, 57, 88, 89, 129, 132, 144,
 242, 334
 negation (denial) 34, 35, 41, 44, 52, 76, 126,
 167, 196, 200, 203, 254, 356, 404, 424,
 739, 740, 743, 744, 748–9, 751, 753
 negative assertions 34–5, 36, 37, 42, 43, 44,
 45–6
 negative existentials 107, 108, 109–10, 112–13,
 346, 351, 363
 negative facts 163, 164, 166, 167, 173, 254
 negative truths, truth-makers for 345–7, 348,
 349
 Nehamas, A. 20 n. 22
 network analysis 552–3
 Neurath, O. 219, 227, 230, 300
 neutralism about the future 443, 445
 Nicholas of Autrecourt 59 n. 37
 nihilism 503, 504, 513, 514, 515, 543
 Niiniluoto, I. 636, 640 n. 23
 Nolan, D. 350 n. 34
 nominalism 69, 461–2, 503, 504, 505, 507, 512,
 513, 514, 651, 659
 non-classical theories of truth 739–54
 noncognitivism 579, 580–1, 582, 584, 688
 and Frege-Geach problem 588–92, 598
 and theory of truth-aptness 584–6
 see also expressivism
 noncontractive approaches 750–2, 753
 non-existence *see* not/non-being

- nonfactualism 503, 535 n. 20
 non-identity 245, 251
 nonindexical contextualism 524
 no-no paradox 679–81, 682, 684
 nontransitive approaches 749–50, 753
 not/non-being 27–32, 40, 41, 43, 46, 76, 77,
 109, 110, 111, 112, 113, 126, 243

 objective facts 94
 objective falsehoods 153, 157
 objective reality 79, 433
 objective representations 102, 104
 objective truth 98
 objectivity 4, 93, 96, 114, 118, 120, 433–7, 456
 in mathematics 648–9, 658–64
 in science 608, 609–10
 of truth-bearers 101–6, 114
 of truth-values 436
 objects 94, 95, 101, 104, 109, 118, 239–40, 247,
 255, 278, 309, 719
 appearances of 97–101, 105, 106
 bundle theory of 141
 cognition and its 95–8, 102
 complexes as 169
 existence or non-existence of 109–10, 113
 and judgment, harmony between 110
 mathematical 462, 466–8
 physical 434, 463–4, 465, 466, 544,
 545, 546
 property instantiation 571
 truth-values as 204–8
 objectual truth 40–1, 42–3
 observation 226–7, 228, 229, 235, 289
 observational acceptances 232, 233
 observation sentences 300
 Ockham's razor 352–3, 447
 Ogden, C. K. 297, 580 n. 1
 Oliver, A. 388
 ontological pluralism 569–70, 572
 ontological realism 435, 462, 463–5, 468
 ontological truth 14, 16, 23, 24, 52, 53
 ontologico-epistemic concept of truth 13, 15,
 16–17, 20
 ordinals 414
 originals 20, 21
 oscillation 757, 763
 Owen, G. 31

 Paoli, F. 752
 paracomplete theories 746, 747, 748–9, 750,
 753, 756
 paraconsistent theories 746–7, 748–9, 750, 753
 paradigms, scientific 605–6
 paradox 5–6, 154
 Burali-Forti 414
 and classical logic 742–5
 Curry's 677, 727, 748, 749
 lottery 645
 Paradox of Sharp Boundaries 421
 preface 645, 675–6
 Russell's 781
 Sorites (Paradox of the Heap) 420–1,
 428–30
 see also semantic paradox
 Parallel Postulate 415, 416
 paraphrase fictionalism 506–7
 Parfit, D. 673
 Parmenides 16–17, 18
 Parsons, C. 771, 775, 777, 782 n. 60
 Parsons, J. 346
 Parsons, T. 264 n. 14
 past 232, 461
 pathological sentences 765, 778
 Patterson, D. 512
 Peano arithmetic (PA) 413, 656, 663, 711, 720,
 723–5, 727, 728, 729, 736
 Pedersen, N. J. L. L. 555 n. 22, 559 nn. 34–6,
 562 n. 39, 571–2 nn. 55–7
 Peirce, C. S. 283, 284–90, 291, 293, 297, 298,
 299, 301, 302, 422, 611, 612
 perception 111, 118, 169, 178–9
 perfection 23–4
 personal supposition 70
 perspectives, science/scientific knowledge
 609, 610
 Pettigrew, R. 657
 Pettit, P. 531 n. 16
phantasia 33 n. 66
 phenomenalism 465–6
 phenomenal world 17, 18
 phenomenology 94
 of ideas 79–80
 physical facts 457
 physical objects 299, 434, 463–4, 465, 466,
 544, 545, 546

- physicists 17, 18
 picture theory of propositions 151, 171–2,
 184–5
 plain truths 557, 562, 563
 platonism 435
 Plato 2, 9–10, 11, 15, 16, 18, 60, 140, 676
 judgments 24, 25–6, 27, 32
 propositional falsehood 26–32
 propositional truth 24–32
 theory of Forms 15, 18–24, 25, 51
 Cratylus 19 n. 20, 21 n. 25, 26, 28, 33 n. 67
 Euthydemus 27 n. 49, 28, 29
 Meno 24 n. 42, 25 n. 43
 Parmenides 20 n. 23, 22, 26 n. 46
 Phaedo 15, 18 n. 19, 19 n. 20, 21 n. 27, 22,
 24 nn. 41, 42
 Phaedrus 15, 18 n. 19, 19 n. 20
 Philebus 15, 18 n. 19, 21 n. 27, 24 n. 42, 26–7
 Politics 18 n. 19
 Republic 15, 18 n. 19, 19 n. 20, 21, 22, 23, 24
 n. 41, 25, 26 n. 46, 27 n. 49, 51
 Sophist 21 n. 25, 22, 26, 27–8, 31, 32 nn. 63,
 65, 34, 35, 36, 45
 Symposium 15, 19 n. 20, 21 nn. 25, 31, 22,
 26 n. 46
 Theaetetus 18 n. 19, 21 nn. 25, 31, 26, 27, 32,
 517, 675
 Timaeus 15, 18 n. 19, 22, 24 n. 41, 25 n. 43,
 26 n. 45
 platonism 462, 466, 467
 mathematical 648–9, 658, 659–61
 pluralism
 logical 570
 ontological 569–70, 570–2
 semantic 569
 see also truth pluralism
 polysemy pluralism 554, 555
 Popper, K. 626, 627–8, 629, 631, 635, 643, 669,
 670
 positive facts 173
 possibilities 187, 272, 307
 possible truth-teller 679
 possible worlds 272, 313, 316 n. 27, 318, 319,
 320, 330, 773
 fictionalism about 506–7, 515
 relativism and 521
 possible world semantics 314–15
 practical truth 33
 pragmatism 3, 283–303, 611
 Dewey 283, 284, 293–7, 299, 300, 302
 James 283, 284, 290–3, 294, 295, 298, 299,
 300–1, 302
 Peirce 283, 284–90, 291, 293, 297, 298, 299,
 301, 302
 Quine 299–300
 Ramsey 176, 183–9, 190, 284, 297–9, 302
 Rorty 284, 300–1, 302
 Pratt, J. 301
 Prawitz, D. 372, 373–4, 383, 385, 392, 655
 predicate-representation 102, 104
 predicates 34, 88, 90, 170, 239, 240, 248, 249,
 255, 278, 310, 311, 484
 predication 323–4, 327–9
 predicative judgments 27, 32
 pre-established harmony 89
 preface paradox 645, 675–6
 prefix fictionalism 505–8, 509, 510–11, 512,
 514, 515
 pre-judgmental truth 33
 pre-propositional truth 33
 presentism 351
 pretense fictionalism 507, 508, 509, 510,
 511–12, 514, 515
 Prior, A. 261, 395, 681
 prisoner's dilemma 673
 probability 626, 644, 645
 Proof of Cases 403
 proof-likeness of truth-makers 358,
 359–60
 proofs, mathematical 649–50, 655
 proof theory 360, 367 n. 20
 proper names 204, 205, 206, 208, 255, 308,
 309, 310, 311, 317–18, 442, 448
 propositional attitudes 307, 322, 325–6, 692
 propositional constituency 316
 propositional falsehood 26–32
 propositional functions 198
 propositional relation 321, 322–3
 propositional truth 10, 11–12, 14, 51, 53–6
 Plato 24–32
 propositions 4, 70, 75–6, 84, 86–7, 102–6, 110,
 115, 118, 119, 125, 126, 127, 132–3, 136,
 139–40, 221, 241, 307–32, 720
 algebraic accounts of 316–17, 330

- atomic 163, 164, 165, 167 n. 12, 171, 173, 571,
633, 635, 720
- compound *see* compound propositions
- disjunctive 164
- elementary 184, 185, 187
- existential 141–2
- Fregean *see* Fregean propositions
(thoughts)
- imperative 327
- as incomplete symbols 156, 160, 161, 173,
176, 177
- instantiation of 325
- interrogative 327
- meanings of 173, 174–5
- Mill 238, 249 n. 15
- picture theory of 151, 171–2, 184–5, 190
- possible world approach to 314–15, 316
n. 27, 319, 320, 330
- real 55–6
- representational capacities 321, 323
- Russell/Russellian 147–8, 152, 153, 155, 156,
158–9, 160, 161, 162, 173, 174, 177–8, 179,
221, 251–4, 269–71, 272, 311–13
- senses of 173, 174–5
- singular 251
- spoken/verbal 55, 85
- and states of affairs, identity between 271–3
- structured or neo-Russellian account of
317–20, 321–3, 330
- structure of 252, 256
- symbolic 147–8
- as truth-bearers 240, 247, 249–50, 251,
259–60, 307
- truth conditions 321, 322–3, 326, 327–8, 330
- unity of 75, 159, 256, 271, 273, 312–13
- Wittgenstein 162–76
- written 55
see also false propositions; true propositions
- Protagoras 17, 38 n. 77, 676
“Man is the Measure” doctrine 17 n. 16,
517, 675
- pure truth content principle 628
- Putnam, H. 219, 225, 233, 357, 410 n. 10, 436
n. 2, 448, 451 n. 15, 615, 656, 698 n. 10
- quantification 484
- substitutional 481, 482, 484, 487, 488
- quantifier domains 771–4, 775, 777, 781–2
- quantum logic 427, 428–9
- quietism 274–7, 279
- Quine, W. V. 225, 234, 299–300, 503, 684, 688
- Rami, A. 355, 360, 364 n. 13, 375
- Ramsey, F. P. 3, 151–2, 161, 166, 176–89, 190,
284, 297–9, 302, 696, 697
- multiple-relation theory 177–80, 181
- pragmatist theory of content 176, 177,
183–9
- and redundancy theory 176, 177, 181–3, 190,
220
- “Critical Notice” of Wittgenstein 182, 184,
185, 188
- “Facts and Propositions” 150, 151, 161, 176,
177, 178, 182–3, 185, 187, 188, 189, 190
- “General Propositions and Causality” 298
- “Truth and Probability” 298
- Ray, G. 705 n. 32
- Read, S. 677, 678
- real distinction 66, 67
- real experience 289
- realism 4, 433–4, 503
- and bivalence 439–40, 441, 461
- common-sense 463–4
- entity 615–16
- mathematical 658–64
- modal 444
- ontological 435, 462, 463–5, 468
- and scientific knowledge 603, 613, 614–16,
620, 621, 622
- semantic 435–46, 449, 450, 461, 468
- structural 616
- and truth-maker theory 349, 373, 389–91
- and truth pluralism 544–6
- reality 9, 13–14, 125, 127, 128–9, 134, 135, 142,
143, 209, 220, 221, 278, 433–4
- coherence theory of 221, 222, 223, 224, 229
- cumulative 232
- degrees of 22–3
- and intellect 54, 56, 62
- objective 79, 433
- true 16, 17–18
- and true propositions 268–9
- qualities 130
- quandaries 459–60

- real propositions 55–6
 reasoning 16
 recognitional capacity 436
 recognition transcendence 436, 437, 438, 439,
 441, 443, 444, 448–9, 451, 456–7, 461–2,
 463, 464, 468
 recollection 60
 rectitude (rightness) 52
 reductionism 461
 redundancy theory of truth 176, 177, 181–3,
 184, 190, 220
 reference 140, 141, 210, 211, 247, 266, 268, 276,
 308, 309, 310, 440, 460
 causal/historical theory of 448–9
 direct 318–19
 of sentences, truth-value thesis 195–7,
 203–4, 212, 309
 regress argument 129–30, 132, 144, 208–12,
 262 n. 9, 266 n. 22
 rehabilitation 757, 763, 769, 772, 775
 Reinhardt, W. 734 n. 19
 relation expressions 308, 309
 relations 126–9, 143, 260, 278, 311–13, 316
 external 131, 138, 139, 143–4, 152
 internal 127, 130–2, 138, 143–4, 145, 359
 relating relation 146
 relation in itself 146
 unreality of 126, 127, 129, 130, 132, 133
 see also multiple-relation theory
 relative truth 5, 17 n. 16, 17, 22 n. 32, 519–26
 relativism 231–2, 434
 assertion 534, 535
 assessment sensitivity 523–5
 belief 535
 centered worlds 525–6
 content 521–2, 524
 context and circumstance 521–2
 deontic modals 527
 disagreement 529–34, 535–8
 eavesdropping 531–2, 537–8
 epistemic modals 526, 527, 528
 future contingents 527–8
 indexical 524
 knowledge ascriptions 527
 moral 517–19
 possible worlds 521
 predicates of taste 520–1, 522, 526, 527, 528, 530
 problems for 534–9
 and scientific knowledge 608, 610
 as self-refuting 534
 and subjectivity 528–9
 traditional 517–19
 release 742, 745, 750, 752
 religion 290–1
 repetition 757, 759–63, 769, 772, 775
 representational truth 12
 representation(s) 93, 94, 95, 96, 97, 98, 99,
 100, 102, 104–5, 321, 322, 324
 and judgments 107, 108–9
 objective 102, 104
 predicate 102, 104
 subjective 102, 104
 Restall, G. 570, 740
 revenge liar, contextual accounts of 755–7,
 759–83
 revolutionary fictionalism 505, 506, 507,
 508–9, 510, 511, 512, 514, 515
 Richard, M. 320, 520, 527, 531 n. 16, 533–4
 Richards, I. A. 580 n. 1
 rigid designators 318–19, 338
 Ripley, D. 749, 750
 Robinson, A. 709 n. 46
 robots, and liar paradox 671
 Rorty, R. 284, 300–1, 302
 Rosen, G. 505, 506, 510, 515
 Rumfit, I. 184, 187
 Russell, B. 3, 90, 91, 102, 109, 110, 125, 127,
 130–1, 136, 138, 144, 145–8, 165, 174, 177,
 180, 198, 221, 257, 260, 261, 442, 445,
 671, 674, 679–80, 696
 complexes, theory of 151, 164, 168–70, 172,
 174, 178, 179, 189
 correspondence theory of truth 147–8, 150,
 151, 155, 159–60, 161–2, 163, 168, 172, 175,
 182, 189, 190, 238, 239, 244, 251–3
 criticism of James 292
 judgment, theory of 150–1, 152–62, 168–9,
 170–1, 189
 multiple-relation theory 127, 145, 146–7,
 148, 151, 154–62, 172, 176, 178, 189, 190
 propositions 147–8, 152, 153, 155, 156, 158–9,
 160, 161, 173, 174, 177–8, 179, 221, 251–4,
 269–71, 272, 311–13, 330
 substitution, theory of 160

- "The Classification of Relations" 146
 "The Monistic Theory of Truth" 154
 "The Nature of Truth" 154, 156, 157, 160
 "On Denoting" 150
 "On the Nature of Truth and Falsehood"
 150, 154, 156, 169
 "The Philosophy of Logical Atomism"
 147–8, 157
The Principles of Mathematics 127, 145, 146,
 150, 159, 311
 "Theory of Knowledge" 146, 160
 Russell's paradox 781
 Ryle, G. 684

 Sainsbury, M. 422–3
 Sainsbury, R. M. 680, 681
 salience structure 773–4, 778
 Salmon, N. 269, 270, 320
 Santas, G. 23 n. 39
 satisfaction 290, 294–5, 357, 706, 707–8, 723
 saturated expressions 308, 309
 saturated formulae 358, 361–2, 377, 380
 with a connective dominant 368–9
 primitive 367–8
 with a quantifier dominant 369
 saturated literals 379, 380
 Schafer, J. 531
 Scharp, K. 757 n. 2
 schematic reflective closure 733
 Schlick, M. 226–7, 228, 230
 Schröder, E. 201, 202
 Schroeder, M. 596 n. 5, 599 n. 6
 Schurz, G. 632 n. 14, 633, 634, 637, 643
 science/scientific knowledge 56–9, 63, 87,
 296, 299, 472–3, 494
 abstract models 619–20
 anti-realist positions 614–15, 616–17
 and approximate truth 614, 620–1, 622
 and bias 609–10
 and coherence theory 604, 606, 608,
 614–15
 and constructive empiricism 613
 convergence 610–12
 and correspondence theory 604, 606, 607,
 609–10, 611, 616, 618, 621
 and empiricism 603, 609, 613, 614
 evolutionary view of 611–12
 feminist critiques of 604, 608–10, 618
 historicist approaches to 604–6, 611–12
 idealized models 619, 620
 incommensurability of paradigms 605–6
 inference to the best explanation 617, 618
 instrumentalist view of 614, 615 n. 7, 621, 622
 and meaning 606, 608
 normal science periods 605–6, 615
 objectivity 608, 609–10
 and observables/unobservables 614–15
 optimism regarding truth and 602–3
 perspectives 609, 610
 pessimism regarding truth and 603
 pessimistic induction argument 618
 Plato's notion of 19
 realist view of 603, 613, 614–16
 and relativism 608, 610
 and scientific models 618–22
 scientific revolutions 605
 skepticism 603–4, 616–18
 social constructivist view of 604, 607, 608
 sociology of (SSK) 606–8, 610, 618
 truth as aim of 610–14
 truth as an achievement of 614–18
 truth as orthogonal to 604–10
 underdetermination of theory by data 617–18
 world change phenomenon 605, 606
 scientific method 286
 scientific realism 603, 613, 615–16, 620, 621, 622
 Searle, J. 589 n. 2, 590
 second-order functionalism 556, 560, 562
 second-order logic 661–2
 self-defeat 670, 672–4, 675, 676, 687, 690
 self-locating contents 525–6
 self-refutation 534, 672
 see also self-defeat
 semantic anti-realism 373, 457, 458
 manifestation argument for 446–52
 semantic-descent deflationism 481, 483–4,
 485, 486, 487, 488, 494, 498, 500
 semantic metaphysics 435–74
 semantic paradox 194, 512–14, 669–94, 736
 contextual approach to 755–86
 no-no paradox 679–81, 682, 684
 and truth pluralism 547
 truth-teller paradox 678–9, 684, 689
 see also liar paradox

- semantic pluralism 569
- semantic realism 435–46, 449, 450, 461, 468
 Dummett 438–46, 460
 Wright 456–60
- semantic truth 41, 42, 43, 44, 45, 68–72
- semantics 193–5, 224, 247, 510, 722, 723
 classical 440
 descriptive 448
 exclusionary 184, 187–8
 foundational 448
 index 314
 mathematical language 653–8
 possible world 314–15
 success 184, 185–7
 of vague statements 422–7
- semblances 20–1, 24 n. 42
- sensations 98, 222, 234
- sense experience 60, 61, 64, 289
- sense(s) 266, 275, 276, 308, 309, 310, 311
 complex 309
 determinacy of 398–403
 unsaturated 309, 310
- sensory cognition 62, 64
- sentences 241, 248
 atomic 253, 254, 309, 310, 356, 440, 441, 720
 declarative 33–5
 mental 69, 247, 248, 250, 252
 observation 300
 structure of 252
 true 250
 as truth-bearers 240, 246, 247, 254–5
 truth-values as referents of 195–7, 203–4, 212, 309
 see also propositions
- sentential relations 321–2, 323
- set theory 399, 663, 722
 failures of bivalence in 410–14
 Zermelo-Fraenkel 408, 411, 412, 413, 415, 416, 417, 418, 651, 663
- Shapiro, S. 458, 557 n. 30, 661
- Sheard, M. 730
- Sher, G. 544 n. 2, 555–6 n. 23
- Sidner, C. 759, 760
- signification 54–6, 69, 70
- simile of the sun 15, 23, 24 n. 40
- Simmons, K. 769 nn. 31, 32
- Simons, P. 345, 355
- simple supposition 70, 71
- singularity theory 767–70, 778, 779–81
- singular propositions 251
- situations 775–7, 778
- skepticism 59
 and scientific knowledge 603, 616–18
- slingshot argument 255–6
- Slote, M. 264 n. 15, 265 n. 16
- Smith, B. 355
- Soames, S. 269, 270, 320, 321, 323–5, 329, 330, 698 n. 10, 700 nn. 16, 17, 708, 715
- social constructivism 604, 607, 608
- sociology of scientific knowledge (SSK) 606–8, 618
 Strong Program (Edinburgh School) 607–8, 610
- Solomon, G. 713
- sophists 17
- Sorites Paradox 420–1, 428–30
- soul 24
- Spade, P. V. 677
- Speaks, J. 325–6, 329, 330
- Spinoza, B. de 76, 79, 80–3, 85, 86, 88, 89, 95, 221, 223
- spoken propositions 55
- Stalnaker, R. 272, 307, 315, 316 n. 27, 525, 534, 757–8, 759
- standpoint theory 609, 610
- Stanley, J. 535, 762
- Stanovich, K. 674
- statements 24, 25, 268
 decidable and undecidable 439, 450
 as truth-bearers 240
- states of affairs 40, 42, 45, 56, 110, 165, 171, 253, 334, 444–5, 478
 Brentano 101, 114, 117
 Husserl 114, 115, 116, 118, 119
 Moore 141, 143
 non-obtaining 242, 243, 244, 252, 272, 275
 as possible worlds 272
 and propositions, identity between 271–3
 and situations 775–7
 as truth-makers 364 n. 13
 Wittgenstein 243, 443
- Stephenson, T. 520, 527, 531 n. 16
- Stevenson, C. L. 590
- Strawson, P. F. 231, 261, 267–8, 449, 450

- strong truth pluralism 553–5, 566–9
 and mixed discourse 568–9
 response to instability challenge and unity
 challenge 567–8
- Stroop effect 672
- structuralism
 mathematical 661
 modal 656–7, 659
- structural realism 616
- structure 246, 248, 249
 and truth-bearers 249–53
- structured propositions 317–20, 321–3, 330
- Stumpf, C. 110
- subatomism 254–5
- subjective representations 102, 104
- subjectivity 93, 120
 and relativism 528–9
 transcendental 119–20
- subjects 34, 88, 90, 239, 240, 248, 249, 278, 311
- subjunctive conditionals 465, 466
- substantivalism 478, 488, 489–93, 496,
 499–500
- substitutional quantification 481, 482, 484,
 487, 488
- substitution theory 160
- substructural approaches 748–52
- success semantics 184, 185–7
- Sundell, T. 539
- Sundholm, G. 358
- superassertibility 455–6, 471, 472, 545–6, 556,
 557, 570
- supervaluationism 444, 445
- supposition 69–72
 co-supposition 71–2
 material 70
 personal 70
 simple 70, 71
- Swyneshed, R. 677
- sylogistic theory 35, 54
- symbolic propositions 147–8
- syntax 321–2, 325
- T-schema 357, 478, 479, 480, 481–3, 484–5,
 487, 488, 494, 513, 514, 515, 599–600,
 726, 727, 728, 729, 741–2, 745, 747, 752
- T-sentences 698–701, 704–5, 715
- Talisse, R. 294
- Tallant, J. 351, 352
- Tappolet, C. 548 n. 9, 550 n. 12
- Tarski, A. 6, 36–7, 193, 255, 357, 362 n. 10, 395–6,
 440, 485, 513, 685–7, 688, 695–717, 721,
 723, 741, 765, 780
 Convention T 697, 698, 699 n. 14, 700–1,
 704, 705–6, 715, 720, 721, 722
 and correspondence theory 696
 essential richness thesis 708, 713–14
 incoherence view of truth 705, 706
 and inconsistency theory of truth 512
 indefinability theorem 396, 397, 703, 704,
 708, 709, 712–13
 inexpressibility theorem (Tarski's theorem)
 708, 709–12
 and liar paradox 686, 687, 696, 698, 702–4,
 706
 logical consequence 371, 372
 truth definition 696, 697–8, 701, 706–8,
 719
 T-sentences 698–701, 704–5, 715
 “The Concept of Truth in Formalized
 Languages” 193, 194
- Taschek, W. W. 458
- taste
 contextualist view 530, 531, 539
 relativist view 520–1, 522, 526, 527, 528
- temporal facts 12, 16
- temporality 16, 22, 232
 of truth value 37–40
- term formalism 656 n. 15
- terminism 50, 68
- that-clauses 262–4, 307, 315
- things-in-themselves 100, 105, 106, 128, 222,
 223, 234, 235
- Thomas Aquinas *see* Aquinas
- thought experiments 289, 296
- thought(s) 143, 239, 240, 250, 266
 discursive 125, 126–7, 129, 140
 Fregean *see* Fregean propositions
 (thoughts)
 non-discursive 127
- three-value logic 688–9
- Tichý, P. 628, 632, 634, 640 n. 23, 641, 643
- “to be (*einai*)” 11, 11–12 nn. 5, 6, 13–14, 30, 36
- transcendental idealism *see* Kant
- transcendentals 51–2

- transcendental subjectivity 119–20
- transparency 741, 742, 745, 752
- true beliefs 152, 153, 154, 186, 287, 291, 292, 293, 295, 301, 307
- true ideas 66, 76, 78, 79–80, 81, 82, 83, 84
- true judgments 155, 156, 157, 159, 181, 239, 239–40
- true propositions 103, 104, 141, 142, 153, 155, 158, 175, 221, 250, 252, 312
- and facts, identity between 260–79
- and reality 268–9
- and states of affairs, identity between 272
- true reality 16, 17–18
- true sentences 259
- true/truth distinction 64–5
- truth-apptitude 470, 471, 473, 551, 584–6, 587–8
- truth-ascription deflationism 488–93, 496, 497, 498, 500
- truth-bearers 4, 11, 14, 39–40, 45, 94, 239, 240–1, 242, 246, 254–5, 259, 262, 276, 277, 278, 279
- assertions as 240
- basic 241, 246
- beliefs as 240
- judgments as 240
- objectivity of 101–6, 114
- primary 241, 246, 247
- propositions as 240, 247, 249–50, 251, 259–60, 307
- secondary 241
- sentences as 240, 246, 247, 254–5
- statements as 240
- and structure 249–53
- truth conditions 164, 165, 166, 168, 170, 171, 172, 175, 181, 435, 441, 443, 446, 448–52, 469, 506–7
- decidable statements 450
- domain of 773–4, 778
- mathematical statements 467–8
- propositions 321, 322–3, 326, 327–8, 330
- recognition transcendent 448–9, 451, 463
- undecidable statements 450
- truth definition, Tarskian 696, 697–8, 701, 706–8, 719
- truthfulness 14
- truth-functions, sentential connectives as 200–4
- truthlikeness 5, 620, 625–47, 670
- content approach 628, 629–31, 643
- likeness approach 628, 634–7, 643
- complexity problem 639–41
- extension problem 637–9
- language-dependence problem 641–2
- relevant consequence approach 628, 632–4, 643
- extension problem 634
- worthlessness problem 634
- see also* approximate truth
- truth-makers/truth-making 4, 40–4, 46, 94, 110, 259, 262, 268, 276, 277, 278, 279, 333–54, 355–93
- and brute truths 337, 338, 339, 345, 350–2
- and causation 333–4, 335
- and conjunctions 375
- and essentialism 334
- formal co-inductive definition of 379–81
- as internal relations 359
- and *in virtue of* relation 335, 336–9
- language-independence 392
- and logical consequence 359, 360, 371–5, 383–6
- M*-relative truth-makers 365–6
- negative truths 345–7, 348, 349
- proof-likeness 358, 359–60
- and realism 349, 373, 389–91
- reflexive instability problem 370–1
- states of affairs as truth-makers 364 n. 13
- and theory of truth 348
- transformation by valid argument 376–9
- truth-maker maximalism 341–2, 345–7, 352
- truth-maker necessitarianism 334, 335, 339–45, 388
- truth pluralism 478, 498–9, 543–75
- double-counting objection 551–2, 566
- Game Analogy Argument 546
- instability challenge 548, 549 n. 10, 560–1, 567–8
- and mixed discourse 549–51, 561–5, 568–9
- principle-based 552–3, 555, 561
- and realism/anti-realism debate 544–6
- and semantic paradox 547
- unity challenge 548–9, 561, 567–8
- see also* moderate truth pluralism; strong truth pluralism
- truth-properties 478, 487, 490, 491, 493, 496, 497, 498–9, 500

- truth-status 435, 439, 443
truth-tables 200, 201, 594, 595, 731
 see also Kleene schema
truth-teller paradox 678–9, 684, 689
truth-value realism 649, 658–64
truth-values 200, 201, 230, 241, 435, 440
 concepts as functions from objects to 195,
 197–200
 objectivity of 436
 as objects 204–8
 as referents of sentences 195–7, 203–4,
 212, 309
 temporal character of 37–40
 truth-value gaps 439, 442, 443, 444, 677,
 690, 731
Turing, A. 685
typed truth theories 723–5
type-free truth 725–36
 classical and symmetric theories 728–30
 disquotational theories of 726–8
 Friedman-Sheard (FS) theory 730, 731
 Kripke-Feferman (KF) theory 730, 732–4,
 734 n. 19
 Non-classical truth with truth-gaps 730–4
 theories with the domain of significance
 734–6
undecidable statements 439, 450
universal facts 254
univocal strong pluralism 554–5
unplain truths 557, 562, 563
unsaturated expressions 308–9, 310
Uzquiano, G. 414
vague statements 444–5, 457, 626
 bivalence 419–22, 430
 epistemicism 444
 semantics and logic of 422–7
 Sorites Paradox (Paradox of the Heap)
 420–1, 428–30
 supervaluationism 444, 445
value of content for falsehoods (VCF) 628,
 630–1, 639
value of content for truths (VCT) 627, 629,
 630, 635, 643
value-ranges of functions 206–7
Velleman, D. 518–19, 540
Venn, J. 126, 201
veridical being 11–12, 17 n. 16, 36, 40, 41, 42,
 43, 44, 46
verification, and vague statements 424, 425, 426
verificationism 219, 227–30, 235, 580
verisimilitude 620, 669, 670, 673, 689–90
Vlastos, G. 22–3 n. 36, 25 n. 43
Walton, K. 511
warrant 373
 assertibility 132–3, 230–1, 294, 453–6, 470,
 471, 546
Weatherson, B. 524, 527
Weidemann, H. 39 n. 80
Weingartner, P. 632 n. 14, 633, 634, 637, 643
Weir, A. 749
Welchman, J. 296
Weston, T. 416, 635 n. 18
Whitehead, A. N. 674
White, M. 299
Wiggins, D. 285, 286, 287
will, and intellect 83
William of Ockham 57–8, 64, 68–72
Williamson, T. 261, 397, 401–2, 429, 473
Wittgenstein, L. 3, 147, 151, 152, 160 n. 9, 162–76,
 177, 180, 181, 188–9, 189–90, 201 n. 14,
 234, 251, 253, 608
 and correspondence theory of truth 151,
 162–5, 167, 172, 190, 243
 on facts 163, 164, 165–76
 multiple-relation theory 172, 184, 190
 picture theory 151, 171–2, 184–5, 190
 Blue Book 164
 Notes on Logic 170, 173
 Tractatus 150, 151, 162–5, 167, 169, 170, 171,
 172, 175, 183, 190, 243, 443
Wolff, C. 95
Woodruff, P. 25 n. 43
Woozley, A. D. 261
Wright, C. 421–2, 435, 436, 437, 438, 441, 446,
 447–8 n. 12, 449–51, 473, 545 n. 4, 546
 nn. 5, 6, 547 n. 7, 553, 554, 555 n. 22, 556
 n. 26, 559 nn. 34–6
 critique of deflationism 452–6
 minimalist theory 437, 452, 454–6, 470,
 471, 546 n. 5
 semantic criteria for realism 456–9

written propositions 55

Xeniades 17

Xenophanes 15, 669, 670, 673, 689,
690

Yablo, S. 507, 508, 512, 515, 557, 657

Yablo's paradox 682–3

Young, J. O. 226

Zalta, E. 316

Zardini, E. 752

Zeno 17

Zermelo, E. F. F. 411–12

Zermelo-Fraenkel set theory with Choice
(ZFC) m 663 , 408, 411, 412, 413, 415,
416, 417, 418, 651

Zimmerman, A. 535

Zwart, S. D. 642–3

