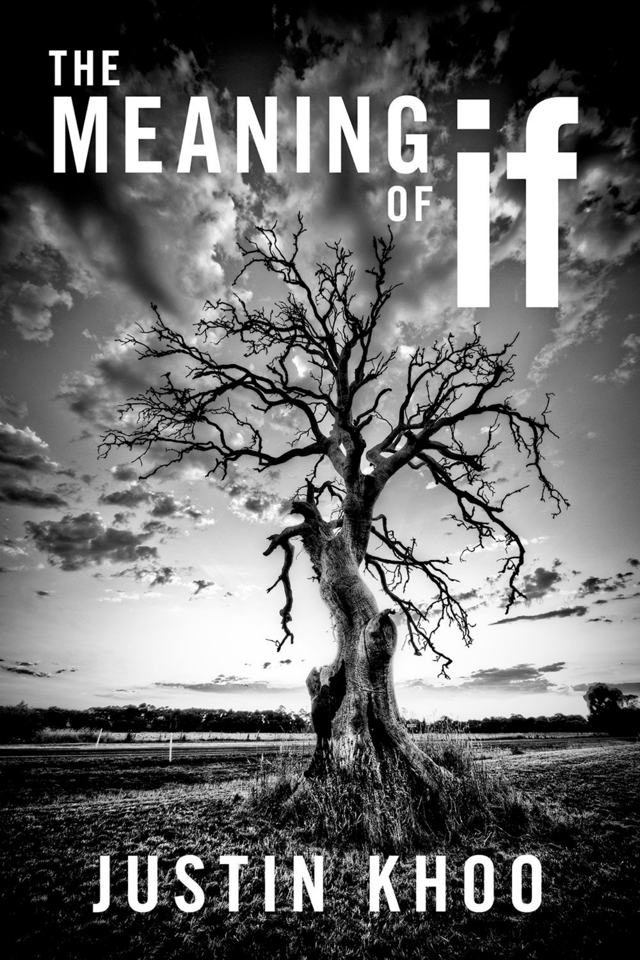

THE MEANING OF if
JUSTIN KHOO

The Meaning of *If*

The Meaning of *If*

JUSTIN KHOO

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

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.

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

CIP data is on file at the Library of Congress.

ISBN 978–0–19–009670–0

DOI: 10.1093/oso/9780190096700.001.0001

Printed by Integrated Books International, United States of America

The manufacturer's authorized representative in the EU for product safety is
Oxford University Press España S.A. of Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en or product.safety@oup.com).
OUP España S.A. also acts as importer into Spain of products made by the manufacturer.

To E.M.K., may you find wonder in the mundane and overlooked.

Contents

Preface

This book grew out of my dissertation, supervised by Zoltán Gendler Szabó at Yale, and the main ideas have slowly been developed and refined during the past seven years at MIT. Most of the material is new, though the seeds of some of the ideas can be found in previous work: for instance, Chapter 6 grew out of my paper "Probabilities of Conditionals in Context" (Khoo 2016), Chapter 7 is a refinement of my thesis in "On Indicative and Subjunctive Conditionals" (Khoo 2015), and Chapter 8 a reworking of the central ideas in "Backtracking Counterfactuals Revisited" (Khoo 2017). It's also worth mentioning that, although I set out to write my panoptic thesis on conditionals—encompassing everything I had to say on the topic—sympathy for the reader led me to abandon that goal. In particular, what you have before you is primarily a motivation for two main ideas: one, that conditionals encode inferential dispositions, and two, that indicative and subjunctive conditionals share a common semantic core. Focusing on these ideas, which I think will be of broad interest to both philosophers and linguists, necessitated setting aside numerous interesting complications about the compositional semantics of conditionals that I would have otherwise loved to explore. Alas, one only has so many pages.

I mentioned refinements, and there is no way this project would exist in the form that it does without the input and commentary of many many friends and colleagues, among them: David Balcarras, Chris Barker, David Boylan, Alex Byrne, Fabrizio Cariani, Sam Carter, Nate Charlow, Adam Cooper, Keith DeRose, Josh Dever, Alexander Dinges, Cian Dorr, Kevin Dorst, Vera Flocke, Kai von Fintel, Tyrus Fisher, Branden Fitelson, Kelly Gaus, Dan Greco, Caspar Hare, Dan Harris, Sally Haslanger, John Hawthorne, Irene Heim, Ben Holguín, Larry Horn, Sabine Iatridou, Stefan Kaufmann, Harvey Lederman, Rose Lenehan, Hanti Lin, Matt Mandelkern, Vann McGee, Eliot Michaelson, Matt Moss, Sarah Moss, Bryan Pickel, Brian Rabern, Agustín Rayo, Daniel Rothschild, Bernhard Salow, Paolo Santorio, Tamar Schapiro, Miriam Schoenfield, Ginger Schultheis, Nathaniel Schwartz, Adam Sennet, Kieran Setiya, Brett Sherman, Patrick Skeels, Brad Skow, Jack Spencer, Bob Stalnaker, Una Stojnic, Megan Stotts, Eric Swanson, Zoltán Gendler Szabó, Judy Thomson, Patrick Todd, Elmar Unnsteinsson, Ian Wells, Roger White, Steve Yablo, and Seth Yalcin.

If I may just shout out a few people (at the risk of overlooking the contributions of others, to whom I preemptively apologize): thanks especially to David Boylan and Ginger Schultheis for reading the entire manuscript and providing detailed comments that led to several significant refinements; thanks to Paolo for being such an excellent interlocutor and coteacher during NASSLLI 2018, where we taught Probabilities of Conditionals in Modal Semantics and pushed each other to better understand the triviality literature; thanks to Daniel Rothschild and Jack Spencer for helping me think through and refine my ideas about the bounding puzzles when they were in their infancy; thanks to my coteachers at MIT—Steve, Vann, Kai, and Sabine—each of our seminars provided fruitful opportunities to try out (and often discard) new and weird ideas, some of which made it into the book; and finally, thanks to Zoltán, who taught me how to be a philosopher, and continues to teach me how to be an advisor.

Finally, and most importantly, thank you to my parents Robert and Sheryl, and my wife Laura. You supported me through thick and thin, and afforded me the chance to pursue this wild and weird project, which must look from the outside like the Mentaculus.[1] I promise you it is not (at least, I think it isn't).

[1] See *A Serious Man* (2009), dir. Ethan Cohen and Joel Cohen.

If you meet a snake,
that is good, it will die and you'll go to Heaven.
If the snake meets you,
that is bad, you will die and the snake will go to Heaven.

—Marin Sorescu, *Omens*
(Trans. Michael Hamburger)

Introduction

I am not one for manifestos, but they can be useful in motivating a project, providing a ladder for us to later throw away after reaching a vantage from which to glimpse a puzzle tantalizing enough to demand our attention for its own sake. One influential foundational manifesto (at least for the tradition of analytic philosophy within which this book is situated) comes from Wilfred Sellars:

> "To achieve success in philosophy would be, to use a contemporary turn of phrase, to 'know one's way around' with respect to all these things [cabbages and kings, numbers and duties, possibilities and finger snaps, aesthetic experience and death], not in that unreflective way in which the centipede of the story knew its way around before it faced the question, 'how do I walk?', but in that reflective way which means that no intellectual holds are barred." (Sellars 1962: 1)

Just as the project of understanding how cabbages and kings hang together with numbers and duties is distinctively philosophical, so is the project of understanding what it is to inquire and reason—to seek knowledge of one's way around things, so to speak. And inquiry and reasoning are most often joint activities, where we turn to language to coordinate our thoughts, and few classes of linguistic expressions are more central to reasoning and inquiry than conditionals:

(1) If it's raining in Dubuque, it is snowing in Denver.

(2) If it hadn't been snowing in Denver, it wouldn't have been raining in Dubuque.

The indicative conditional (1) guides my inquiry into the actual weather patterns: by accepting it, I am prepared to infer that it's

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0001

snowing in Denver upon learning that it's raining in Dubuque. By contrast, the subjunctive conditional (2) encodes thoughts about explanation and dependence: by accepting it, I endorse the thought that rain in Dubuque somehow depends on snow in Denver.[1]

It seems, then, that reflective knowledge of inquiry will involve exploring how conditionals facilitate joint inquiry. But is there really anything puzzling here? We jointly inquire by pooling our information, thus winnowing the range of possibilities compatible with our evidence, and we often do this by talking. So what? The issue is that the role conditionals play in all this is somewhat unusual. Suppose I believe (and have good evidence for thinking, etc.) that it's raining in Dubuque and I want us to rule out any possibilities to the contrary. A good move would be to assertively utter the sentence:

(3) It's raining in Dubuque.

Why? Because (3) encodes the information I believe and want you to believe, and to assert is to propose that we jointly believe what's asserted (cf. Stalnaker 1978, Lewis 1980a). But now suppose I haven't ruled out the possibility that it is raining in Dubuque, I believe that either it's not raining in Dubuque or it is snowing in Denver, and I want to communicate this belief to you. Then, a good move would be to assertively utter the indicative conditional (1):

(1) If it's raining in Dubuque, it is snowing in Denver.

Assuming you don't have more information about the weather than me, then, upon hearing me say this, you should come to believe exactly what I aim to communicate: that either it's not raining in Dubuque or is snowing in Denver. But if that is right, then we have

[1] We also find ourselves reaching for conditionals when jointly planning (if this happens, then do . . .), explaining (if this had happened instead, . . .), and deciding (if I did this, then . . .). Furthermore, conditionals also underlie attitudes of regret (thinking things would have been better if . . .), relief (thinking things would have been worse if . . .), and fear (worrying that if such and such happens, . . .). Thus, reflective, non-centipedian knowledge of these activities and attitudes will, it seems, require knowing what these conditionals mean and how they connect to these activities and attitudes.

an argument for one of the most vilified positions in the literature on conditionals: the dreaded material conditional theory (according to which the indicative conditional $A \rightarrow B$ is true iff either A is false or B is true). The argument goes as follows:

(i) Accepting $A \rightarrow B$ just is to learn that either A is false or B is true.
(ii) Accepting a sentence just is to believe that it is true.

One reason to worry about the conclusion of this argument is that intuitions about the probabilities of conditionals suggest that indicative conditionals are stronger than material conditionals—that is, they are true in fewer places than the latter. Suppose we know Jones had a fair coin but only think it's 1/2 likely that he flipped it. Since we know the coin is fair, it seems we should be certain that the following is false:

(4) If Jones flipped the coin, it landed heads and tails.

But if indicative conditionals are equivalent to material conditionals, then in fact we should think that (4) is 1/2 likely, equal to the probability that Jones didn't flip the coin plus the probability that he flipped it and it landed heads and tails.

Already, we can see that inquiring with conditionals raises a host of issues we don't find with ordinary declarative sentences. And the problem isn't limited to indicative conditionals. Suppose that you know Jones didn't flip the coin, but are not sure whether his coin was double-headed, double-tailed, or fair (suppose you regard each possibility as equally likely). A trusted informant tells you:

(5) If Jones had flipped the coin, it would have landed heads.

Accepting this claim, you should rule out the double-tailed possibility, but also rule out the possibility that the coin was fair.[2] After all, if you knew it was fair, you would know that:

[2] There is a backtracking interpretation of (5) which is such that learning that would not be enough to rule out any property of the coin. It's paraphrased roughly as follows:

(i) Jones would have flipped the coin only if it had been heads. So, if he had flipped the coin, it would have landed heads.

(6) It might be the case that if Jones had flipped the coin, it would have landed tails.

but then if you could also know (5), you would then know the epistemic contradiction:

(7) #If Jones had flipped the coin, it would have landed heads and it might be the case that if Jones had flipped the coin, it would have landed tails.

but this seems like something you couldn't come to know, or even rationally believe (cf. Yalcin 2007). So, learning (5) seems sufficient for learning that it is necessarily true (given the kind of coin that Jones had) that either he didn't flip the coin or that he flipped it and it landed heads. Plugging this result into our argument above leads to the conclusion that the subjunctive conditional $A \Rightarrow B$ is true if and only if it is necessarily true (given the kind of coin that Jones had) that either he didn't flip the coin or that he flipped it and it landed heads.

However, this conclusion is implausible. Initially, when you regard it as equally likely that Jones's coin was fair, double-headed, or double-tailed, you should think that the conditional (5) is $1/2$ likely to be true. This is because you regard it as $1/3$ likely that the coin is double-headed, in which case (5) would be certainly true, and $1/3$ likely that the coin is double-tailed, in which case (5) would be certainly false, and $1/3$ likely that the coin is fair, in which case (5) would be $1/2$ likely. The problem is that the predicted truth conditions for (5) are too strong: it is necessary (given the kind of coin that Jones had) that either he didn't flip the coin or that he flipped it and it landed heads only at worlds where Jones's coin is double-headed. Thus, our conclusion leads to thinking that (5) can be at most $1/3$ likely to be true (equal to the probability that the coin is double-headed).

We have two puzzles, motivated by commonplace reflections on our use of conditionals in joint inquiry. I call these bounding puzzles because they reveal a tension in the strength of conditionals—material

Set this interpretation aside. I will come back to discuss backtracking interpretations in Chapter 9.

conditionals seem to put an upper bound on the strength of indicative conditionals, a bound they seem desperate to break free from, and strict conditionals seem to put a lower bound on the strength of subjunctive conditionals, yet they seem weaker nonetheless.

The philosophical literature on conditionals is riddled with puzzles like these, prompting some rather extreme responses. Among them are views that flatfootedly accept either the material or strict conditional theory—such positions go back to Philo and Diodorus (see Kneale & Kneale 1962). Someone defending such a view might take the intuitions about what we learn from conditionals as motivation and attempt to explain away the boundary-busting truth-conditional intuitions to the contrary. More recently, the modal theories of Stalnaker 1968, Lewis 1973a aim to account for the truth-conditional strength/weakness of indicative and subjunctive conditionals, but lack an account of why what we learn from conditionals diverges from their truth conditional content. Even more recently is the trend to treat (some) conditionals as truth valueless, lacking propositional content altogether—instead, holding that accepting a conditional is a matter of your belief state having a certain global property (Edgington 1995; Bennett 2003; Yalcin 2011, 2012a; Moss 2015). Perhaps, in the case of indicative $A \rightarrow B$, this property is being disposed to infer B upon learning A. However, theories in this nonfactualist tradition face trouble accounting for the fact that we seem to assign non-trivial probabilities to conditional contents, suggesting that they can be more or less likely true. After all, what would it mean to assign a probability to a property of your cognitive state?[3]

It seems, then, that we need a new way of thinking about conditionals, their contents, and the role they play in our cognitive lives. This book is my attempt to provide such a theory.

[3] Philosophical hand-wringing over conditionals isn't limited to the Western tradition either. We find discussions of the logic of conditionals in Ancient Chinese philosophy of logic (especially in the Moist school; see Liu & Zhang 2010); and Naiyyāyika scholars in Classical Indian Philosophy explored connections between subjunctive conditionals and causality (see Chakrabarti 2010).

Part I: Framework

The book divides into three parts. In Part I, I spell out the two bounding puzzles more precisely, and argue for a framework for thinking about conditional contents that allows us to solve both. The view I defend is situated in a modal framework similar to that developed by Stalnaker, and shares features with the material, strict, and non-factualist theories sketched above. In particular, the truth value of a conditional does not supervene on the facts—instead, it draws non-factual distinctions between possibilities representing differences in inferential dispositions. Conditional contents encode inferential dispositions—dispositions to infer some factual content from learning some other factual content. I argue that what it is to believe, think possible, and invest some probability in a conditional content is a matter of your inferential dispositions, and your inferential dispositions are grounded in your factual beliefs. This allows me to define the modal properties of conditional contents (the conditions under which they are necessarily, possibly, probably true, etc.) in terms of modal properties of (and relations between) non-conditional, factual contents. Conditional contents can be genuinely believed and thought more or less probable, even though they do not have factual truth conditions.

While conditional contents are (sometimes) non-factual, information never is. This motivates distinguishing the informational content of a sentence from its (possibly non-factual) content; in the case of conditionals, I use this wedge to motivate an account of why the information we learn from a conditional diverges from its (non-factual) content—the informational content being weaker in the indicative case, and stronger in the subjunctive case, thus resolving our two bounding puzzles.

Part II: Probabilities

I close Part I with a promissory note: there is a way to assign probabilities to non-factual extended possibilities that will allow us to assign intuitively plausible probabilities to conditional contents. But immediately we face a problem: various triviality results seem to show that this simply cannot be done (see, for instance, Lewis 1976). In Part II, I take up this challenge and argue that my theory can avoid such triviality challenges—indeed, I go further and establish a tenability result, proving that the theory entails a plausible generalization about the probabilities of conditionals.

Still, problems remain. One of my results is a partial vindication of Stalnaker's Thesis—the claim that the probability of a conditional is equal to the conditional probability of its consequent given its antecedent. However, there are unusual scenarios in which this thesis is violated (see McGee 2000; Kaufmann 2004; Moss 2015, 2018; Khoo 2016). I argue that such scenarios motivate relativizing conditional contents to partitions, which may be set by questions under discussion. A pragmatic theory of this partition parameter explains why such interpretations arise and why they are typically quite rare.

Part III: Subjunctive vs. Indicative

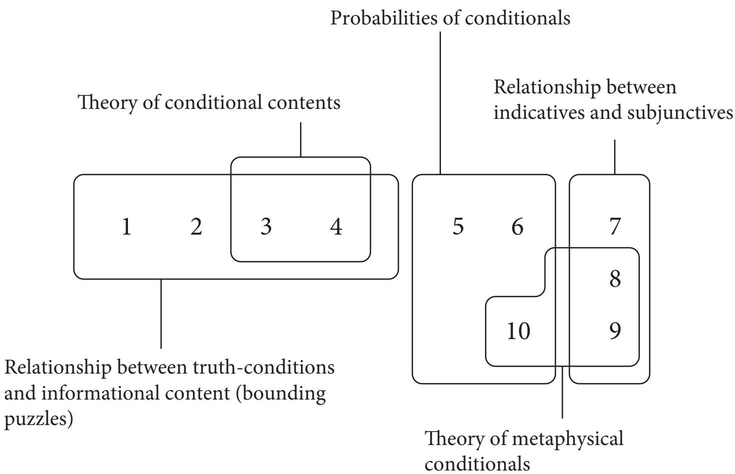
A second issue left open in my discussion of the bounding puzzles is what accounts for the semantic and pragmatic differences between indicative and subjunctive conditionals.

Building on work by Kratzer 1981, 1991, 2012; Iatridou 2000; Condoravdi 2002, I defend a unified theory of indicative and subjunctive conditionals on which their grammatical differences (in particular, the presence of the past perfect *had* and tense auxiliary *would*) drive their semantic differences (both truth conditional and informational), as well as their pragmatic differences (resulting in indicatives presupposing the epistemic possibility of their antecedents while subjunctives implicating that their antecedents are false). Mine is a theory on which the past + future tense morphology on subjunctive conditionals is real, and applies to the modal element in the conditional, shifting the time at which its domain is determined to the past (thus opening up past historical possibilities that would otherwise be closed due to the passage of time, similar to ideas proposed by Thomason & Gupta 1980, Tedeschi 1981, Ippolito 2003, 2006, Arregui 2007).

But why would facts about what goes on at alternative histories matter to causality, explanation, and rational decision making? I argue that not every alternative history matters, but, rather, only those that are like ours with respect to their sufficiency networks—sets of relational properties between true propositions at a world. I use "sufficiency" here broadly to encompass causal, as well as explanatory, and perhaps grounding, sufficiency. Drawing on insights of causal model theories of subjunctive conditionals due to Pearl 2000, Hiddleston 2005, Briggs 2012, I argue that subjunctives encode information about local interventions into a sufficiency network that generates certain downstream changes, which provides information about how propositions are related to one another within the network. I argue that we need to distinguish the intervention point (or time) of the conditional from the time of its antecedent and consequent, which allows us to generate backtracking and non-backtracking interpretations of a single subjunctive conditional, each of which encodes different properties about the network.

Finally, I show how the groundwork laid in Parts I and II extends naturally to subjunctive conditionals, allowing us to predict plausible probabilities for subjunctive conditionals while avoiding some recent triviality results that seem to undermine the possibility of such theories.

Here is a visual layout of the central topics of the book broken down by chapter:



The puzzles we will discuss in this book cut deep to foundational issues in the philosophy of language, but can now be stated precisely with technical tools developed in logic and linguistic semantics. As such, progress on these problems now requires the input of "persistently reflective specialists" (to use Sellars's phrase)—that is, philosophers versed in linguistic semantics, and linguists versed in philosophy of language. I hope this book can make a small contribution to the community of persistently reflective specialists thinking hard about conditionals, and also encourage others to come along for the ride.

PART I
FRAMEWORK

1
Bounding Puzzles

In this introductory chapter, I want to introduce the puzzle that will animate the first part of this book. The puzzle stems from two natural ways of thinking about the meanings of sentences that pull in opposite directions when it comes to conditionals.

The first way of thinking about meaning relates sentences to the world by way of their truth conditions (necessary and sufficient conditions for its truth). Whatever else we say about meaning, it seems that the meaning of a declarative sentence is, or at least determines, the conditions under which it is true or false—its truth conditions. A sentence's truth conditions play a role in delineating its logical properties (what it entails and is entailed by), as well as how likely it is (its likelihood of truth), and the conditions under which we may utter it (perhaps only if we know it to be true). These properties of declarative sentences are fixed by their meaning, and thus it seems one cannot say what a declarative sentence means without articulating (in some way) its truth conditions.

The second way of thinking about meaning relates sentences to mental states by way of their acceptance conditions. The meaning of a sentence determines the conditions under which it is correctly accepted by an individual—this may describe a certain state of belief, or perhaps an affective or conative state. A sentence's acceptance conditions play a role in our rational evaluation of speakers, as well as in communication, since in speaking one aims (often and among other things) to coordinate certain properties of one's state of mind with one's interlocutors.

These two thoughts about meaning are unified on standard possible worlds semantic theories, where a possible world is a maximal way things could be—a theoretical object that decides the truth value of every proposition. Possible worlds are useful tools for modeling mental states. For instance, we can think of a belief as a mental state

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0002

that represents things as being a certain way—in other words, it is compatible with some possible worlds and not others. A belief, then, can be thought of as a set of possible worlds, namely, those that are compatible with it.

Just as a belief can be thought of as a set of possible worlds, so too can a sentence's meaning—the set of worlds at which the sentence is true. This is sometimes called a possible worlds proposition, although we'll just call it a **proposition** (ignoring issues of idealization for now). If we do this, then we can identify the conditions under which a sentence is genuinely accepted (its acceptance conditions) with the conditions under which the proposition expressed by the sentence is believed:

> **Acceptance**
> To accept a sentence *A* just is to believe the proposition expressed by *A*.

It follows that to come to accept *A* just is to update your belief state by intersecting it with the set of worlds at which *A* is true. This is the dynamic analog of acceptance or belief—the process of coming to accept or believe. The fact that the theory captures both dimensions of meaning simultaneously is a major theoretical advantage of possible worlds semantics, which may go some way towards accounting for its enduring explanatory success in linguistic semantics (see, for instance, Stalnaker 1978, 1998; Lewis 1980a; von Fintel & Heim 2012).

In this chapter, I argue that conditionals raise a challenge to this simple theory of meaning, since their acceptance conditions come apart from their truth conditions. The puzzle arises for both indicative and subjunctive conditionals, but in different ways.

1.1 Indicative Bounding

Indicative conditionals seem truth conditionally stronger than material conditionals, and yet in many cases it seems that accepting an indicative conditional just is to come to believe the corresponding material conditional.

Suppose you have no idea where Smith is. I tell you:

(1) If Smith is not in Athens, he is in Barcelona.

What do you thereby learn? Plausibly, you learn at least that Smith is not somewhere other than Athens or Barcelona—that is, enough to rule out that Smith is in Chicago, Des Moines, etc. If you did not now rule out such possibilities, then you could believe or say the following, which seems incoherent:

(2) #I agree that if Smith is not in Athens, he is in Barcelona; and maybe he is in Chicago.

This suggests that indicative conditionals obey modus ponens:

Modus Ponens
$A, A \to B \vDash B$

(Here, '$\vDash$' denotes truth conditional entailment—if A and $A \to B$ are both true, then B must also be true.) But it also seems that you learn nothing more than this. For, if you did, what would this additional information be? There are two cases in which you might learn something minimally more than that Smith is in Athens or Barcelona. Case one is that you learn that either Smith is in Athens and X or Smith is Barcelona, while case two is that you learn that Smith is in Athens or Smith is Barcelona and X, for some additional X. Yet, what could this additional information, X, be? I have no clue. Yet, if there really were such a piece of information, we would expect it to be part of the meaning of *if*, and hence detectable by ordinary English speakers—otherwise, how could we reliably communicate it using conditionals? Yet, I submit, there simply is no such additional information communicated by (1).

The material conditional $A \supset B$ is true if its antecedent is false or consequent is true:

A	B	$A \supset B$
T	T	T
T	F	F
F	T	T
F	F	T

Where $A \to B$ is some indicative conditional, we thus have:[1]

Weak Sufficiency

Coming to accept $A \to B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.

Weak Sufficiency suggests that indicative conditionals just are material conditionals, since their acceptance conditionals are almost the same. It's puzzling, then, that indicative conditionals also seem to be truth conditionally stronger than material conditionals. Intuitively, invalid inferences like (3) and (4) are predicted to be valid on the material conditional theory:

(3) Obama isn't now in space. So if he is in space, he's sitting on the moon.

(4) It's raining hard. So if it's not raining, it's raining hard.

Furthermore, the probability of an indicative conditional often seems less than its material conditional counterpart. Suppose I roll a fair die and keep the result hidden. We wonder how likely it is that:

(5) If the die landed on a prime, it landed on an even.

[1] The restriction to scenarios in which you regard $A \wedge B$ as possible is crucial for avoiding the following case:

(i) a. If Sue is in France, she is not in Paris.
  b. If Sue is in Paris, she is not in France.

Suppose I am not sure where Sue is, and then I accept (i-a). It seems that the propositional information I gain is just that either Sue is not in France or she is not in Paris, and thus I learn that Sue is not in Paris (wherever she is). But suppose I am not sure where Sue is—still, I should not accept (i-b), since Paris is in France. But if accepting (i-b) just is to come to believe that either Sue is not in Paris or not in France, this just is the same information I would get by accepting (i-a). The restriction to agents who believe $A \wedge B$ is possible rules this out because in the first case, I believe that it is possible that Sue is in France and not in Paris (maybe she is in Lyons). But in the second case, I do not believe it is possible that Sue is in Paris and not in France.

Intuitively, the probability of this conditional is 1/3—there are three possible prime outcomes (2, 3, 5), one of which is even. Thus, the intuition here is evidence that the probability of (5) just is the conditional probability of its consequent (that the die landed on an even) given its antecedent (that the die landed on a prime).

However, the material conditional theory predicts that the probability of (5) is 2/3. This is because the probability of the material conditional *prime* $\supset$ *even* is equal to the probability of either not rolling a prime or rolling an even. There are three non-prime outcomes (1, 4, 6) and three even outcomes (2, 4, 6), and thus four of six possible outcomes satisfy this condition: 1, 2, 4, 6.

In fact, in general, the probability of $A \supset B$ will be greater than the conditional probability of B given A—of course, except when both equal 1. So, it seems, the material conditional theory is stuck predicting the wrong probabilities of conditionals. This result corroborates the standard thought that indicative conditionals are truth conditionally stronger than material conditionals—that is, the former are true in fewer places than the latter.

We can generalize as follows:

Strength

It is possible that $A \supset B$ is true and $A \rightarrow B$ is false.

However, now we face a problem. Weak Sufficiency and Strength, together with the following principle about learning and conditionalization, entail a contradiction:

Conditionalization$^+$

Coming to believe A is at least to rule out all possibilities in which A is false.

From Strength, there are possibilities in which $A \supset B$ is true and $A \rightarrow B$ is not true. Then, there may be an agent S who does not rule out some such possibilities. By Conditionalization$^+$, if S were to come to believe $A \rightarrow B$, she must rule out those possibilities, but that would be to rule out some possibilities in which $A \supset B$ is true, and hence come to believe something more than that $A \supset B$ is true, contradicting Weak Sufficiency.

1.2 Subjunctive Bounding

In the previous section, we looked at indicative conditionals like (1):

(1) If Smith is not in Athens, he is in Barcelona.

For our purposes (for now—this will be refined and explained in more detail in later chapters), the subjunctive counterpart of the indicative (1) is:

(6) If Smith hadn't been in Athens, he would have been in Barcelona.

Subjunctive conditionals are distinguished from indicatives both in their grammar and what they mean—grammatically, by the presence of extra tense morphology, and in what they mean by the observation that subjunctives seem to be about objective relations in the world rather than epistemic relations between propositions (more on both distinctions in Chapter 7).

Subjunctives are subject to their own kind of bounding puzzle, which can be put basically like this: subjunctive conditionals seem truth conditionally weaker than what I will call strong subjunctive conditionals, and yet in many cases it seems that one fully accepts a subjunctive conditional only if they believe the corresponding strong subjunctive conditional.[2]

Alice had a coin, which was either fair, double-headed, or double-tailed. Let's stipulate that if the coin is fair, then it could have been the case that Alice flipped heads and it could have been the case that Alice flipped tails (the modals here understood metaphysically, which

[2] This is closely related to a puzzle developed by Schulz 2017; however, I articulate it in terms of learning, which reveals the assumption Conditionalization^{-}. I suspect, though, that Schulz and I may ultimately agree on the resolution to this puzzle. See Chapter 3 and Khoo 2020 for a discussion of how my strategy and his differ.

means that these possibilities are at least compatible with knowing that Alice didn't flip the coin); if the coin is double-headed, Alice only could have flipped heads; and if the coin is double-tailed, Alice only could have flipped tails. You are sure she did not flip it, but then you come to accept (perhaps via an assertion by a trusted informant):

(7) If Alice had flipped the coin, it would have landed heads.

Accepting this requires ruling out the double-tailed possibility. But it seems that accepting this also requires ruling out the fair coin possibility. After all, if you do not rule out this possibility, it seems you should regard it as an open possibility that Alice could have flipped tails, in which case you should think it is an open possibility that:

(8) If Alice had flipped the coin, it would not have landed heads.

But then it seems you would be committed to an epistemic contradiction (to use Yalcin 2007's terminology):[3]

(9) #If Alice had flipped the coin, it would have landed heads, and maybe if Alice had flipped the coin, it would not have landed heads.

The infelicity of (9) suggests that either the agent doesn't really accept (7) or else has ruled out the possibility of (8). This is similar to Yalcin's diagnosis of simple epistemic contradictions like:

(10) #It's raining and it might not be raining.

[3] We could put this point another way as well (cf. Santorio 2021). If Alice could have flipped tails, it seems true that if Alice had flipped the coin, it might not have landed heads. But then, having learned (7) without ruling out this possibility, you should be able to learn:

(i) #If Alice had flipped the coin, it would have landed heads, but it might not have landed heads if Alice had flipped it.

Yet, this seems like something you could not rationally learn.

Yalcin proposes that the infelicity of (10) is due to the fact that ϕ informationally (though not truth conditionally) entails $\neg\Diamond\neg\phi$, and this means, roughly, that accepting the former requires accepting the latter. Turning back to our case, ruling out the fair possibility leaves you with only the double-headed possibility, relative to which the conditional is true.

Let's define a **strong subjunctive conditional** roughly as follows: $A \boxright B$ is true iff A couldn't have been the case without B also being the case. This is rough, since we haven't yet specified the relevant domain of the possibility modal *could have*, but I hope this at least gives us a sense of things.[4] Thus, at a world where her coin is double-headed, it couldn't have been the case that Alice flips the coin and it lands on anything other than heads, and likewise for worlds where the coin is double-tailed. But at worlds where the coin is fair, it could have been the case that Alice flips heads and could have been the case that Alice flips tails. Thus, the strong subjunctive $F \boxright H$ is true only at the double-heads worlds. Thus, our observation above was this: you accept that had Alice flipped the coin, it would have landed heads only if you believe the strong subjunctive $F \boxright H$.

Where $A \Rightarrow B$ is an arbitrary subjunctive conditional, our observation motivates the following general principle:

> **Strong Sufficiency**
>
> If you are sure that A is false, then you accept $A \Rightarrow B$ only if you believe $A \boxright B$.

However, there is also compelling evidence that subjunctive conditionals are weaker than strong subjunctive conditionals. Suppose you know Jones has a fair coin and did not flip it. Then, the subjunctive conditional:

(11) If Jones had flipped the coin, it would have landed heads.

[4] As a preview, I will define the truth conditions of a subjunctive conditional relative to a domain of possible worlds D, and then the strong subjunctive will be defined so that all A-worlds in D are B-worlds.

should intuitively have probability 1/2. But the corresponding strong subjunctive conditional $F \mathrel{\Box\!\!\rightarrow} H$ has probability 0, since you are certain that it is not the case that Jones could only have flipped heads. Generalizing, we have:

Weakness

It is possible that $A \Rightarrow B$ is true and $A \mathrel{\Box\!\!\rightarrow} B$ is false.

Again, both principles seem plausible, but, together with the following principle about learning and conditionalization, entail a contradiction:

Conditionalization^{-}

To come to believe A is at most to rule out all possibilities in which A is false.

From Weakness, there are possibilities in which $A \Rightarrow B$ is true and $A \mathrel{\Box\!\!\rightarrow} B$ is false. Then, there may be an agent S who does not rule out some such possibilities. By Conditionalization^{-}, if S were to come to believe $A \Rightarrow B$, she must not rule out those possibilities, but that would mean she would believe $A \Rightarrow B$ without believing $A \mathrel{\Box\!\!\rightarrow} B$, contradicting Strong Sufficiency.

1.3 Zooming Out

We thus have two sets of three incompatible principles. Again, the puzzle can be put simply: intuitions seem to pull the truth and acceptance conditions of conditionals apart. Indicative conditionals seem truth conditionally stronger than material conditionals, but the information you learn from an indicative seems to be that of a material conditional:

Strength: $A \supset B \not\models A \rightarrow B$

Weak Sufficiency: Learning $A \rightarrow B$ is to learn no more than $A \supset B$

And subjunctives seem truth conditionally weaker than strong subjunctives, but the information you learn from a subjunctive seems to be that of a strong subjunctive:

Weakness: $A \Rightarrow B \not\models A \boxminus\!\!\rightarrow B$

Strong Sufficiency: Learning $A \Rightarrow B$ requires learning $A \boxminus\!\!\rightarrow B$

How can this be? And why do indicatives and subjunctives differ in this regard?

Here is the space of possible options, if we want to preserve as many of these principles as possible:

Indicative Bounding

	Weak Sufficiency	Strength	Conditionalization^{+}
1.	✓	✓	X
2.	✓	X	✓
3.	X	✓	✓

Subjunctive Bounding

	Strong Sufficiency	Weakness	Conditionalization^{-}
1.	✓	✓	X
2.	✓	X	✓
3.	X	✓	✓

In the first part of this book, I will argue that we should resolve the Indicative Bounding Puzzle by giving up Weak Sufficiency, and that we should resolve the Subjunctive Bounding Puzzle by giving up Conditionalization^{-}. My theory will be presented in Chapter 3. In Chapter 2, I situate the bounding puzzles within the existing literature and articulate some challenges facing existing strategies for handling the puzzles.

2

Situating the Puzzles in the Literature

In this chapter, I will discuss how various views in the literature can be understood to respond to the two bounding puzzles discussed in Chapter 1, and articulate some challenges facing each strategy.

Terminological Disclaimer

I will use uppercase italic letters $A, B, C, \ldots$ as variables over sentences, and uppercase boldface letters **A**, **B**, **C**, ... as variables over propositions. I adopt the following notational conventions:

- $\bar{\mathbf{A}}$ = the set of worlds at which **A** is false.
- **AB** = the set of worlds where **A** is true and **B** is true.

'$A \rightarrow B$' is a sentence schema over indicative conditionals, and '$A \Rightarrow B$' is a sentence schema over subjunctive conditionals (we will distinguish them more carefully in Chapter 7). $[\![\ ]\!]$ is a semantic evaluation function that maps expressions to their contents relative to a context c. Here, I will assume that the semantic content of a sentence relative to a context c is a proposition, and thus that $[\![A]\!]^c = \mathbf{A}$, ignoring relativization to the context unless keeping track of it matters. Given this assumption, the extension of a sentence at a context c and world w just is the truth value (either 1, for true, and 0, for false) it has at c, w: $[\![A]\!]^{c,w} = 1$ iff $[\![A]\!]^c$ is true at w.

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0003

2.1 Indicative Bounding

Start with the Indicative Bounding Puzzle. Remember that the puzzle is that indicative conditionals seem at once stronger than material conditionals (Strength), but also such that believing an indicative conditional does not involve believing anything stronger than the corresponding material conditional (Weak Sufficiency). To avoid the puzzle, we must thus deny either Strength, Weak Sufficiency, or the principle connecting them (Conditionalization$^+$). I discuss each in reverse order.

2.1.1 Denying Conditionalization$^+$

To deny Conditionalization$^+$ would be to hold that one can believe something without ruling out all of the possibilities in which it is false:

> **Conditionalization$^+$**
>
> Coming to believe *A* is at least to rule out all possibilities in which *A* is false.

An immediate problem with this strategy is that it predicts it is possible to believe *A* while at the same time leave open the possibility of ¬*A*. But that seems wrong. Utterances of the following are infelicitous:

(1) a. #It is raining and, moreover, it might not be raining.
 b. #It is raining and, moreover, it is possibly not raining.

Furthermore, if we were to learn that S believes it is raining, it no longer seems to be an open question for us whether she regards it as an open possibility that it is not raining.

This is not to say that this strategy is without some plausibility. As some have noted, belief ascriptions seem weak, in the sense that *S believes that p* is roughly equivalent to *S believes it is likely that p* (cf. Hawthorne et al. 2016; Rothschild 2019). That view predicts that it should be acceptable to say:

(2) I think it is raining, but I have not yet ruled out the possibility that it is not raining.

And, indeed, there is some intuitive pull to think that (2) is felicitous, especially when compared with the sentences in (1).

However, this is cold comfort to someone denying Conditionalization$^+$ as a means of avoiding the Indicative Bounding Puzzle, for we can simply restate the puzzle using a notion of full belief, which requires ruling out all possibilities to the contrary. Indeed, note that Weak Sufficiency seems equally plausible when formulated with a notion of full acceptance:

Weak Sufficiency (Full)

Coming to fully accept $A \rightarrow B$ (when you regard $A \wedge B$ as possible) is to come to (fully) believe nothing more than that $A \supset B$ is true.

Summarizing:

Problems for denying Conditionalization$^+$

While denying Conditionalization$^+$ avoids the Indicative Bounding Puzzle, the puzzle can be restated in terms of full belief in a way that avoids commitment to Conditionalization$^+$.

2.1.2 Denying Weak Sufficiency

Denying Weak Sufficiency amounts to holding that one does learn more from an indicative conditional than its corresponding material conditional:

Weak Sufficiency

Coming to accept $A \rightarrow B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.

A famous theory that generates violations of Weak Sufficiency comes from Robert Stalnaker (Stalnaker 1968, 1980). Stalnaker proposes that conditionals are true relative to a world w and a selection function f, which is a function from a world w, and a proposition $\mathbf{A}$, to a world, $f(w, \mathbf{A})$—the closest $\mathbf{A}$-world to w. Selection functions are defined to obey the following constraints:

Selection functions

(i) *Success*: $f(w, \mathbf{A}) \in \mathbf{A}$.
(ii) *Minimality*: If $w \in \mathbf{A}$, then $f(w, \mathbf{A}) = w$.
(iii) *CSO*: If $f(w, \mathbf{A}) \in \mathbf{B}$ and $f(w, \mathbf{B}) \in \mathbf{A}$, then $f(w, \mathbf{A}) = f(w, \mathbf{B})$.

We can now state Stalnaker's semantics for conditionals as follows:

Stalnaker Semantics
$[\![A \to B]\!]^{w,f} = 1$ iff $f(w, \mathbf{A}) \in \mathbf{B}$.

This says that $A \to B$ is true iff the closest $\mathbf{A}$-world is a $\mathbf{B}$-world. Given the definition of selection functions, it follows that $A \to B \vDash A \supset B$, so indicatives are predicted to be at least as strong as material conditionals. But the theory also predicts that indicative conditionals are truth conditionally stronger than material conditionals (in line with Strength)—that $A \supset B \nvDash A \to B$. To see why, notice that there may be $\bar{\mathbf{A}}$-worlds w such that $f(w, \mathbf{A})$ is a $\bar{\mathbf{B}}$-world, in which case $A \supset B$ is true at w and $A \to B$ is false at w.

Now, what are the acceptance conditions for a conditional, given Stalnaker Semantics? Notice that the semantics does not immediately provide us a proposition that can serve as what we believe when we come to accept the conditional. However, two options exist in the literature. The first is that context c supplies a unique selection function f_c, which generates a unique proposition expressed by $A \to B$ relative to c:

Contextualist Stalnaker Semantics
$[\![A \to B]\!]^{c} = \{w\colon [\![A \to B]\!]^{c,f_c,w} = 1\}$

The second strategy Holds that to accept a conditional requires accepting the proposition it would express relative to every admissible selection function:

> **Strong Stalnaker Semantics**
> To accept $A \to B$ in c just is to believe $\{w: [\![A \to B]\!]^{c,f,w} = 1\}$ for every admissible f.

Both views predict counterexamples to Weak Sufficiency, and thus avoid the indicative bounding puzzle. However, it is unclear what resources they have to explain the intuitive appeal of Weak Sufficiency. According to Contextualist Stalnaker Semantics, accepting $A \to B$ in context c requires eliminating all $\neg A$-worlds w at which $f_c(w, \mathbf{A}) \notin \mathbf{B}$. And according to Strong Stalnaker Semantics, accepting $A \to B$ in c requires eliminating all $\neg A$-worlds w if there is an admissible selection function f such that $f(w, \mathbf{A}) \notin \mathbf{B}$.

Summarizing:

Problems for denying Weak Sufficiency

While denying Weak Sufficiency by adopting Stalnaker Semantics avoids the Indicative Bounding Puzzle, it remains unclear how Stalnaker Semantics (understood as either Contextualist Stalnaker Semantics or Strong Stalnaker Semantics) can capture what seemed so plausible about Weak Sufficiency in the first place.

2.1.3 Denying Strength

We come finally to the last option: deny Strength.

> **Strength**
> It is possible that $A \supset B$ is true and $A \to B$ is false.

The most straightforward denial of Strength comes from the material conditional theory, on which $\mathbf{A} \to \mathbf{B} = \mathbf{A} \supset \mathbf{B}$. However, as we saw in Chapter 1, a standard argument against the material conditional theory comes from the so-called paradoxes of material implication: predicted validities that in fact seem invalid:

(3) Obama isn't now in space. #So if he is in space, he's sitting on the moon.

(4) It's raining hard. #So if it's not raining, it's raining hard.

The standard response to these concerns holds that these conditionals are true but unassertable, which aims to account for the felt invalidity of these argument forms (Jackson 1979, 1987; Grice 1989; Lewis 1976; Rieger 2006). However, even these sophisticated strategies involving assertability won't help us with the second kind of problem facing a material conditional theory.

Notice that $A \supset B$ is false only if A is true and B is false. Now consider an embedded conditional: *I know it is not the case that if A, B*. According to the theory, this can only be true if you know A. But that seems wrong. Suppose you are not sure whether the jewels were stolen by the butler or one of the groundskeepers (the gardener or the roofer). Later, you learn it wasn't the gardener. It seems you could then say:

(5) I do not know whether a groundskeeper stole the jewels, but I do know that it is not the case that if a groundskeeper stole the jewels, it was the gardener.

But it follows on the material conditional theory that (5) entails:

(6) #I do not know whether a groundskeeper stole the jewels, but I do know that a groundskeeper stole the jewels.

Yet (6) is a contradiction. So, according to the material conditional theory, (5) is also a contradiction; yet it is clearly coherent. This problem is not avoided by appealing to the assertability of the embedded conditionals, since the conditional "If a groundskeeper stole the jewels, it was the gardener" is not asserted in (5).[1]

[1] On Jackson's theory, an indicative conditional conventionally implicates that it is robust with respect to its antecedent (meaning that its probability is high and its

It might seem, then, that the strategy of rejecting Strength is in a tough spot. However, there is another, more radical, way to reject Strength. This strategy holds that conditionals do not have truth values at all, and hence, strictly speaking, do not enter into truth-preserving entailment relations and do not have probabilities of truth. I will call this position **radical expressivism**.

This kind of view is endorsed by No Truth Value theorists like Ernest Adams, Dorothy Edgington, and Jonathan Bennett, perhaps surprisingly, since each of these theorists claims to endorse (slightly different versions of) a thesis relating conditional likelihoods to conditional probabilities:

> "The fundamental assumption of this work is: *the probability of an indicative conditional of the form 'if A is the case then B is' is a conditional probability*... the probability of 'if A then B' should equal the ratio of the probability of 'A and B' to the probability of A (ratio of conjunction of antecedent and consequent to antecedent)."
>
> (Adams 1975: 3)

> "Writing '*b*' for 'your degree of belief in', it may be rewritten,
>
> $$b(B\ if\ A) = b(A\&B)/b(A)$$
>
> Call this '**The Thesis**'."
>
> (Edgington 1995: 262–263)

> "To evaluate $A \to C$, I should (1) take the set of probabilities that constitutes my present belief system, and add to it a probability = 1

probability given its antecedent is also high). One might then try to appeal to the fact that conventional implicatures project out of certain embeddings to handle this embedding issue. Just as (i-a) conveys that Smith owns Acme furniture, so does (i-b):

(i) a. Smith, who owns Acme furniture, is a millionaire.
b. I know it's not the case that Smith, who owns Acme furniture, is a millionaire.

Given Jackson's theory, then, we expect the conventional implicature of a conditional to project out of the embedding in (5):

(ii) I know that it's not the case that if a groundskeeper stole the jewels, it was the gardener.

Thus, Jackson's theory predicts that an utterance of (ii) should communicate both that the speaker knows that the conditional is false but also that she thinks its probability is high. Thus, Jackson's theory predicts that (ii) should be incoherent. Yet, intuitively it is not.

> for *A*; (2) allow this addition to influence the rest of the system in the most natural, conservative manner; and then (3) see whether what results from this includes a high probability for C."
>
> (Bennett 2003: 29)

Each of these quotes connects some measure of likelihood of the conditional (perhaps probability, credence, degree of belief in) with a corresponding conditional probability. The central idea goes back to Ramsey, and it appears in its modern form as a conjecture in Stalnaker 1970, so I will call the version involving probabilities Stalnaker's Thesis:

Stalnaker's Thesis

The probability of $A \to B$ just is the probability of B given A, if the probability of A is nonzero.

On its face, Stalnaker's Thesis is quite plausible. Recall the scenario from Chapter 1: I have just rolled a fair die and kept the result hidden. The probability of the conditional (7) seems to be $1/3$, which just is the conditional probability of rolling an even given that I rolled a prime:

(7) If I rolled a prime, I rolled an even.

However, somewhat confusingly, none of the above theorists actually endorses Stalnaker's Thesis. In other places, each of these theorists makes it clear they reject the view that conditionals have genuine probabilities. Adams, for instance, explicitly denies that compounds of conditionals have well-defined probabilities:

> "The author's very tentative opinion on the 'right way out' of the triviality argument is that we should regard the inapplicability of probability to compounds of conditionals as a fundamental limitation of probability, on a par with the inapplicability of truth to simple conditionals." (Adams 1975: 35)

But if there is no probability of $A \to B \wedge X$, where X is some non-conditional sentence, then in what sense could there be a probability of $A \to B$? The probability of a sentence just is the probability that it is true, and this is given by a measure over a Boolean algebra of propositions. A probability function P is a function from propositions in this algebra to values in [1, 0]. Assuming this function assigns a value to $A \to B$ and X, it must be able to assign a value to their conjunction, which is also an element the probability function is defined on (the intersection of the two).[2] Furthermore, a restriction on embedding is not really what matters, since there are triviality results that involve no embedding or complex conditionals (Hájek 1989).[3] Jonathan Bennett is aware of this point, and he responds to Hájek as follows:

> "Hájek's argument requires $A \to C$ to be a free-standing proposition that can be assigned a probability that is to be *matched* by a certain conditional probability, namely [$P(C|A)$]. The proof shows that there cannot be enough conditional probabilities to provide the needed matches. In Adams's theory, the Equation does not assert an equality; rather, it *defines* $P(A \to C)$ as [$P(C|A)$], telling us what a conditional probability is . . . in particular, $P(A \to C)$ is *not* the probability of truth for a proposition $A \to C$." (Bennett 2003: 104)

The point here is that conditionals do not *have* probabilities; rather, they are devices used to *express* our conditional probabilities. As Bennett puts it, "in asserting $A \to C$, a person expresses his high probability for C given A" (Bennett 2003: 106). Alternatively, conditionals allow us to communicate conditional credences or beliefs, but conditional credences or beliefs are not, according to these expressivist theorists, credences or beliefs in any conditional propositions. Edgington says as much, proposing an argument against propositionality directly via a version of the bounding puzzle:

(i) Ruling out $A \land \neg B$ is enough to be certain that $A \to B$.
(ii) It's not necessarily irrational to disbelieve A and disbelieve $A \to B$.

[2] See also Hájek 2014: 238–241 who raises a similar question for Edgington's view.
[3] We will come back to discuss Hájek's result in Chapter 4.

> "We may generalize. Take any proposition. Either it is entailed by $\neg(A\&\neg B)$, or it is not. If it is, it will satisfy (i) but not (ii) (when substituted for 'if A, B'). If it is not, it may satisfy (ii), but cannot satisfy (i). Conditional judgements interpreted according to the Thesis satisfy both (i) and (ii). So they cannot be interpreted as belief in any proposition." (Edgington 1995: 280)

Recently, dynamic semanticists have found common cause with these early No Truth Value theorists. According to dynamic theories, the semantic values of conditionals are update values—functions from information states to information states—rather than propositional contents. Let $|A|$ be the update value assigned to sentence A, and i be some information state (set of worlds). Then, we might assign indicative conditionals the following update value:

Dynamic Conditionals

$$i|A \to B| = \begin{cases} i \cap \mathbf{A} \supset \mathbf{B} & \text{if } i|A| \neq \varnothing \\ \varnothing & \text{otherwise} \end{cases}$$

This update value differs from standard dynamic theories (see, for instance, Gillies 2004, 2009, 2010; Yalcin 2012b, c; Starr 2021), and is instead inspired by Russell & Hawthorne 2016. This has the advantage of predicting Weak Sufficiency (assuming that $i|A| \neq \varnothing$ throughout) in the following sense. An information state i accepts sentence A just if $i|A| = i$. Then, to accept $A \supset B$ (without accepting $\neg A$) will be enough to accept $A \to B$. And to accept $A \to B$ is to accept nothing more than $A \supset B$.

Notice, though, that the update value of a conditional is not something that a probability operator is defined on.[4] As such, defenders of

[4] There are two values one might define using the update value of a conditional that a probability operator would be defined on. The first is the information state as updated by $A \to B$: $i|A \to B|$. However, if this is defined, then it just is the set of $\mathbf{A} \supset \mathbf{B}$-worlds in i, which will have a probability equal to $\mathbf{A} \supset \mathbf{B}$. The second object is the set of worlds in i pointwise updated by $A \to B$:

$$\bigcup\{w \in i : \{w\}|A \to B| = i\}$$

However, this will just be the set of **AB**-worlds in i, so clearly this is too strong to deliver the intuitive probabilities of conditionals (insofar as the probability of $A \to B$ is often greater than the probability of AB).

this kind of theory might count themselves among those who reject Strength (for an explicit articulation of this point, see Yalcin 2018).[5]

However, by rejecting Strength, these theorists owe us an explanation of our intuitive judgments about the probabilities of conditionals. In the next three sections, I consider three possible error theories for the intuitive appeal of Strength. The first strategy holds that such intuitions really track degrees of assertability or acceptability of the conditional (cf. Adams 1965; Jackson 1979, 1987). The second holds that these intuitions can be accounted for by our intuitions about object language probabilified conditionals of the form, $A \rightarrow$ *probably* B or *probably* $A \rightarrow B$ (cf. Kratzer 1986; Starr 2021; Yalcin 2018). And the third holds that there is a distinct notion of probability that applies to conditionals, call it **shrobability** (cf. Charlow 2020). This strategy invalidates certain applications of the law of total probability that involve combinations of the objects of probability with the objects of shrobability.[6]

Assertability

According to this strategy, conditionals have assertabilities, which are values in the unit interval that measure how willing one is to assert it or to accept an assertion of it (cf. Adams 1965; Jackson 1987):

[5] I also count Moss 2015, 2018 among such views. Moss assigns sets of credal states as the contents of conditionals, and such sets do not have probabilities (although they encode constraints on probabilities). Moss shows how to generate reasonable truth conditions for probabilified conditionals like *if A, probably B*, but does not predict the intuitively correct probabilities for the contents of conditionals (I discuss this aspect of her view in the Appendix to this chapter).

[6] I confess that I have trouble locating Adams 1975; Edgington 1995; Bennett 2003 within this taxonomy. I suspect the reason is that each is primarily concerned with the negative thesis—denying that certain laws of probability apply to conditionals. But we still need an error theory of what we are doing when we seem to judge that conditionals themselves have probabilities. So, I think that these theories are incomplete, and that completing them will involve taking up one of these strategies (or some other strategy I have not considered). As we will see in Chapter 3, I endorse some aspects of the NTV theory—in particular, its commitment to non-propositionalism—but I argue that the non-propositional contents of conditional are still objects of probability, and I show how to extend an ordinary probability operator into one that can assign probabilities to such contents.

Assertability

An indicative conditional $A \rightarrow B$ is assertable to a degree equal to $Pr(B|A)$, if $Pr(A) > 0$.

The view says that our intuitions about the probabilities of conditionals really track the degree to which they are assertable. And, since assertability need not obey the probability calculus, we can block attempts to run a revenge version of the triviality results using assertability in place of probability.

But how plausible is Assertability? Consider an example:

(8) If I toss this fair coin, it will land heads.

How assertable is this? According to Assertability, it is 0.5 assertable. But, that seems wrong. Keith DeRose raises this objection:[7]

> "To the extent that I can just intuit the degree to which the conditional is assertable, I would give it a value much lower than 0.5. (Forced to assign a number, I would go for something like 0.06.) After all, it is a fair coin. So I have no idea which of its two sides it will land on if I toss it. I would have to say that I am in no position to assert either that it will land heads if I toss it, or that it will land tails if I toss it. And it does not seem a close call: neither conditional seems close to being half-way assertable." (DeRose 2010: 12)

Lottery cases provide even more striking counterexamples. Suppose I have entered a fair 1000 ticket lottery. I now consider the following conditional:

(9) If the drawing was held, I lost.

My intuition is that (9) is not assertable at all. Maybe it's a bit more assertable than the coin case, but it's nowhere near perfectly assertable. Yet, since $Pr(L|D) = 0.999$, it is predicted to have a very high degree

[7] See also Hájek 2012's bolstering of this objection.

of assertability. If my intuition here is on the right track, this suggests that Assertability is wrong, and thus cannot account for our intuitions about the probabilities of conditionals. After all, the probability of (1) is intuitively 1/2, and the probability of (9) is intuitively 0.999. The conclusion we should draw is that probability and assertability are different things. Thus, we cannot account for our intuitions about probabilities of conditionals by appealing to assertability.

The Restrictor Gambit

According to the theory under consideration, conditionals do not have probabilities. We might think it follows that sentences such as (10-a) and (10-b), which ascribe probabilities to conditionals, must be false:

(10) a. If I rolled a prime, I probably rolled an odd.
b. Probably, if I rolled a prime, I rolled an odd.

On its face, this may seem impossible. The embedded conditional *if I rolled a prime, I rolled an odd* has no probability, so how could (10-a) and (10-b) be true? The trick is to see that it is possible to interpret (10-a) and (10-b) in which they are not attributing a probability to any sentence (or proposition) but rather are stating a conditionalized probability claim.

There are various ways to generate this result. I will focus on a version in which the *if*-clause shifts the information state relative to which its consequent clause is interpreted (cf. Kratzer 1991).[8] The basic idea is that the probability operator embedded in the consequent of the conditional is evaluated relative to an information state that contains the information supplied by the *if*-clause. Thus, where i is an information state and w a possible world, we have

Restrictor Semantics
$[\![A \rightarrow B]\!]^{i,w} = 1$ iff $[\![B]\!]^{c,i^{\mathbf{A}},w} = 1$.
(Where $i^{\mathbf{A}} = i \cap \mathbf{A}$)

[8] This strategy will involve positing a covert modal in the consequent of bare conditionals—see Kratzer 1986. We'll come back to this idea in Chapter 4.

Recall that i is an information state. We can now define a related notion, that of an evidential state e, which is a pair of an information state i and a probability function P normalized on i. The same information state corresponds to multiple distinct evidential states. We allow that evidential modal bases can be shifted, just like information states, as follows:

Shifting
$e^{\mathbf{A}} = \langle i_e^{\mathbf{A}}, P_e^{|\mathbf{A}} \rangle$, if $P_e(\mathbf{A}) > 0$; undefined otherwise.

We can thus restate the Restrictor Semantics (newly formulated relative to evidential states e) as follows (remember, the view will posit a covert modal in the consequent of a conditional when none is present overtly):

Restrictor Semantics
$[\![A \to B]\!]^{c,e,w} = 1$ iff $[\![B]\!]^{c,e^{\mathbf{A}},w} = 1$.

Next, we can give a simple semantics for *probably* (cf. Yalcin 2010):

$[\![Probably\ A]\!]^{c,e,w} = 1$ iff $P_e(\mathbf{A}) > 0.5$

Now, turn back to (10-a). Putting this all together, we predict that $A \to$ *probably B* is a statement of a conditional probability, rather than a statement that ascribes an unconditional probability to a conditional. Here is the derivation of its truth conditions:

$[\![A \to probably\ B]\!]^{c,e,w} = 1$ iff
$[\![probably\ B]\!]^{c,e^{\mathbf{A}},w} = 1$ iff
$P_{e^{\mathbf{A}}}(\mathbf{B}) > 0.5$ iff
$P_e^{|\mathbf{A}}(\mathbf{B}) > 0.5$ iff
$P_e(\mathbf{B}|\mathbf{A}) > 0.5$

Thus, we predict that (10-a) can be true even if it is not the case that the conditional (7) has no probability (this strategy is explicitly endorsed by Yalcin 2018).

(7) If I rolled a prime, I rolled an even.

Suppose for now that a similar strategy can apply to the wide-scope probabilified conditional (10-b). Still, concerns about the strategy remain. One is that it is not sufficient to account for all the relevant intuitions about conditionals and probability (see Charlow 2016 for a clear statement of this challenge). In particular, consider the following dialogue:

(11) A: If I rolled a prime, I rolled an odd.
B: That is likely.

Here, B seems to predicate *likely* of the content of A's assertion. And, intuitively, B's claim is true just if the conditional probability that A rolled an odd given that A rolled a prime is greater than 0.5. However, here, the restrictor strategy seems in trouble: *that* is anaphoric to the content of A's assertion—B's claim is not equivalent to *likely: S*, where *S* is the **sentence** A uttered. We know this because the truth of B's claim concerns what A rolled, not what B rolled. Suppose A rolled a six-sided die, and B rolled a four-sided die. The conditional probability that A rolled an odd given that A rolled a prime is two-thirds. The conditional probability that B rolled an odd given that B rolled a prime is one-half. Clearly, B's claim here is true because it concerns the outcome of A's roll, not B's.[9]

Bifurcation

The third kind of error theory we will consider holds that there is a measure-theoretic quality distinct from probability that applies to conditionals—call it **shrobability**. Thus, we have a probability function P that applies to subsets of W, and a shrobability function S that applies to subsets of W^*, which is the space in which we find conditionals. Both functions obey Totality and Additivity, which

[9] This is not to say that the restrictor gambit is out of options: von Fintel & Gillies 2015; Ciardelli 2021 offer a more complex account of propositional anaphors, whereby they pick up on a higher semantic type. I leave it as a project for future work to explore the prospects of these more advanced versions of the restrictor gambit.

means that the law of total probability (or, rather its equivalent for shrobability) applies to conditionals, for any partition of W^*. However, we block mixtures of the two spaces. In particular, conjunctions of conditionals and non-conditionals have neither probabilities nor shrobabilities. This allows us to block instances of the law of total probability applied to conditionals (since they lack probabilities), as well as to block instances of the law of total shrobability applied to conditionals (since $S(A \rightarrow B|A \supset B)$ is undefined). I am not aware of anyone who has worked out the details of how to use this approach to avoid the triviality results, but see Charlow 2020, for a version of this kind of approach to accounting for intuitions about the probabilities of conditionals.

Because of the ban on assigning measure-theoretic properties to mixtures of elements of the two algebras, this view runs into challenges when it comes to certain conjunctions. In particular, suppose that Smith has just rolled a fair six-sided die. Before the result is revealed, A, B and C are talking about it:

(12) A: If Smith rolled a prime, he rolled an odd.
B: Smith rolled a 4 or 5.
C: It is unlikely that you are both right.

C's claim here seems correct, since the probability that A and B are both right can be no greater than the probability that just B is right, which is 1/3. But this relies on the principle that the probability of a conjunction can be no greater than the probability of either conjunct individually. However, that reasoning only works if conjunctions have probabilities, and the conjunction of A and B's claims has no probability (or shrobability).

The mixing ban also results in problems accounting for comparative probability claims, where what is being compared is a conditional to a non-conditional. Here is an example:

(13) It is more likely that Smith rolled an odd if he rolled a prime than it is that Smith rolled a 4 or 5.

Given that, according to the bifurcation strategy, there is no single measure-theoretic quality assigned to conditionals and non-conditionals, there is no way to make sense of (13). If *more likely* means "having greater probability," then (13) is either false or nonsensical, since the conditional does not have a probability. And if *more likely* means "having greater shrobability," then (13) is likewise either false or nonsensical, since the non-conditional does not have a shrobability. Either way, it seems (13) should not be true. Yet, it seems clearly true.

Summarizing:

Problems for denying Strength

While denying Strength by adopting either a material conditional theory or a radical expressivist theory avoids the Indicative Bounding Puzzle, none of the strategies considered (the assertability, restrictor, or bifurcation gambits) offered a plausible error theory of the probabilistic intuitions supporting Strength.

2.2 Subjunctive Bounding

We turn now to the Subjunctive Bounding Puzzle. The puzzle is that subjunctive conditionals seem truth conditionally weaker than strong subjunctive conditionals (Weakness), even though believing a subjunctive conditional seems to require believing the corresponding strong subjunctive conditional (Strong Sufficiency). To avoid the puzzle, we must thus deny either Weakness, Strong Sufficiency, or the principle connecting them (Conditionalization$^-$).

2.2.1 Denying Conditionalization$^-$

To deny Conditionalization$^-$ would be to hold that sometimes coming to believe A requires ruling out some possibilities you previously regarded as open in which A is true:

Conditionalization^{-}

To come to believe A is at most to rule out all possibilities in which A is false.

This is quite surprising, to say the least. Suppose $A \wedge B$ and $A \wedge \neg B$ are both compatible with your beliefs. Denying Conditionalization^{-} would mean that in some such cases coming to believe A would require also coming to believe $A \wedge \neg B$.

Moore paradoxical sentences seem to be of this form:

(14) It's raining and no one believes it's raining.

Suppose it is compatible with your beliefs that it is raining and also that no one believes it is raining. Then, (14) is compatible with your beliefs. But, supposing positive introspection holds in this instance, then coming to believe that it's raining would mean coming to believe that you believe that it's raining, and thus believe the instance of $A \wedge \neg B$:

(15) It's raining and someone believes it's raining.

While these do seem to be counterexamples to Conditionalization^{-}, the real issue is whether it is plausible that learning subjunctive conditionals admits of an analogous response. If not, then we can just reformulate Conditionalization^{-} to exclude cases of positive introspection and the problem will still remain.

So, the question is whether there is a similar introspection principle that could ground a similar violation of Conditionalization^{-} for subjunctives. And, it is hard to see why there should be such a principle. Of course, we could stipulate one, but that would not provide us what we want, which is a motivated solution to the puzzle. Moritz Schulz suggests a possibility: a subjunctive conditional's truth value depends on an arbitrary selection of an antecedent-world from its domain, and $A \Rightarrow B$ is true iff an arbitrary **A**-world from its domain is a **B**-world. Schulz then suggests that "we can only know an arbitrary F to be G if all Fs are Gs . . . for [otherwise] we cannot know which [F] has been selected" (Schulz 2017: 176). This suggests that there are facts about which F is selected, as if God rolled some dice at the beginning of the universe that determined all the arbitrary selection facts, and these are facts that we simply cannot know.

Obviously, that last comment was meant figuratively. But a real question looms here: What determines the arbitrary selection facts? And what is the nature of our inability to know such facts? Is it due to our limited cognitive capacities, or are they somehow in principle unknowable—does, for instance, God know them? These are tough questions, and ones it would be preferable not to have to answer. Schulz himself recognizes these issues, and leans towards a supervaluational treatment of statements of arbitrary selection. We will return to explore whether such a strategy can ultimately be made to work below (see §3.4.1 and Khoo 2020).

Summarizing:

Problems for denying Conditionalization^{-}

We lack an independent reason to think Conditionalization^{-} should fail; furthermore, we lack an explanation why Conditionalization^{-} seems to fail only in certain instances (for instance, with subjunctive conditionals); finally, denying Conditionalization^{-} is to hold that we cannot learn certain conditional facts on their own, and we have no clear account of why this would be.

2.2.2 Denying Weakness

Denying Weakness is to hold that subjunctive conditionals are at least as strong as their strong subjunctive counterparts:[10]

Weakness

It is possible that $A \Rightarrow B$ is true and $A \mathrel{\Box\!\!\rightarrow} B$ is false.

[10] Recall that the strong conditional $A \mathrel{\Box\!\!\rightarrow} B$ is true iff (roughly) it couldn't be the case that A is true and B is false.

One strategy might be to deny Weakness while attempting to preserve what seemed plausible about it by fiddling with the closeness relation.[11] However, there are reasons to doubt such strategies will succeed.

For one, as Moss 2013: 253 points out, ordinary subjunctives do not pattern as we would expect strong subjunctives to:

(16) I can't say for sure what would happen if I were to flip this coin one million times.

(17) #It's not the case that if I were to flip it one million times, it would land heads each time.

Assuming the coin is fair, it is possible that were one to flip it one million times, it would land heads each time. But the corresponding strong conditional ($MF \boxright H$) is false—thus, if subjunctives are strong subjunctives, we would expect (16) to be false and (17) to be obviously true, yet those judgments are intuitively reversed. Similarly, if we help ourselves to an English paraphrase of $MF \boxright H$ as *if I were to flip this coin one million times, it would definitely land heads each time*, then we get the result we would expect:

(18) It's not the case that if I were to flip it one million times, it would definitely land heads each time.

But the intuitive truth of (18) is then further evidence against thinking that subjunctives are strong subjunctives.

Another problem for this strategy is that subjunctive conditionals (and conditionals in general) seem to validate Conditional Excluded Middle:

Conditional Excluded Middle

$\vDash (A \Rightarrow B) \vee (A \Rightarrow \neg B)$

[11] I have in mind here strategies building on Lewis 1979b's discussion of "quasi miracles" such as Bennett 2003's "near miss" proposal, Williams 2008a's strategy to constrain conditional domains to "typical" worlds, and Lewis 2016's strategy to constrain conditional domains to "relevant" worlds. See Hawthorne 2005 for worries about Bennett's version, worries which both Williams's and Lewis's strategies avoid.

Here is an argument for CEM from Stalnaker 1980. Consider a dialogue involving a universal quantifier:

(19) A: Carter has to appoint a woman to the Supreme Court.
B: Who does he have to appoint?
A_1: He doesn't have to appoint any particular woman; he just has to appoint some woman or other.
A_2: I don't know; I just know it's a woman that he has to appoint.

Here, A_1 is appropriate if A intended *a woman* to scope below *has to*, while A_2 is appropriate if A intended *a woman* to scope above *has to*. Thus, strong modals like *has to* generate scope ambiguities with indefinites, precisely for the same reason that they do not allow negation to "push in":

$$\neg\Box A \nvDash \Box\neg A$$

But conditionals seem to pattern differently—they do not generate a similar ambiguity with indefinites:

(20) A: Carter would have appointed a woman to the Supreme Court if there had been a vacancy.
B: Who would he have appointed?
A_1: #He wouldn't have appointed any particular woman; he just would have appointed some woman or other.
A_2: I don't know; I just know he would have appointed a woman.

Here, A_1's response does not make any sense. This is like saying,

(21) #Carter will appoint a woman to the Supreme Court, but he won't appoint any particular woman.

Thus, like *will*, subjunctive conditionals do not generate scope ambiguities with indefinites, which is evidence that they are not strong modals, and instead validate CEM.[12]

[12] For more arguments in favor of CEM, see §3.1.

But Conditional Excluded Middle and the denial of Weakness do not sit well together. Consider the theory that denies Weakness because it holds that $A \Rightarrow B$ is equivalent to $A \mathrel{\Box\!\!\rightarrow} B$. Assuming it could have been that **AB** and could have been that $\mathbf{A\bar{B}}$, then both $A \Rightarrow B$ and $A \Rightarrow \neg B$ will be false.

We might try to avoid this consequence by proposing that this can never happen by assuming that an utterance of $A \Rightarrow B$ presupposes that either it could only have been that **AB** or it could only have been that $\mathbf{A\bar{B}}$ (cf. von Fintel & Iatridou 2002). However, this strategy then runs afoul of the observation that there are two ways for $A \Rightarrow B$ to be epistemically possible. One is that for it to be epistemically possible that it could only have been that **AB**. But the other way is for it to be epistemically necessary that it could have been that **AB** and could have been that $\mathbf{A\bar{B}}$.

An example of the first case is the following. You think it is possible that Smith had a double-headed coin and also possible that he had a double-tailed coin. On this basis, you think it is possible that:

(22) If Smith had flipped his coin, it would have landed heads.

An example of the second case is the following. You fully believe that Smith had a fair coin, and thus that he could have flipped heads and could have flipped tails—this also seems to be a case in which you think (22) is possible. However, the Weakness-denying theory that tries to rescue Conditional Excluded Middle we are considering cannot predict this. Instead, the theory will predict that the conditional (22) has a false presupposition in this latter case.

Summarizing:

Problems for denying Weakness

The central issue with denying Weakness is that subjunctive conditionals do not behave like strong subjunctive conditionals. Knowing that our coin is fair is sufficient for not knowing either whether it would have landed heads had it been flipped or would have landed tails had it been flipped. Furthermore, Conditional Excluded Middle is invalid for strong subjunctive conditionals but seems valid for subjunctive conditionals.

2.2.3 Denying Strong Sufficiency

To deny Strong Sufficiency while accepting the other principles is to allow that one can learn $A \Rightarrow B$ even though one does not believe $A \mathrel{\square\!\!\rightarrow} B$:

Strong Sufficiency

If you are sure that A is false, then you accept $A \Rightarrow B$ only if you believe $A \mathrel{\square\!\!\rightarrow} B$.

Since we accept Weakness, we hold that $A \Rightarrow B$ is weaker than $A \mathrel{\square\!\!\rightarrow} B$. That means that we could be in the following situation: you are certain that Smith had a fair coin and that he did not flip it, and hence believe that Smith could have flipped the coin and it landed tails (where the *could* here is non-epistemic), and then also believe that nonetheless if Smith had flipped the coin, it would have landed heads. That is quite strange. Notice that assertions of such conjunctions are infelicitous, suggesting an irrationality similar to that of believing that p and might not-p:

(23) #Smith could have flipped the coin and got tails, but if Smith had flipped the coin, it would have landed heads.

Summarizing:

Problems for denying Strong Sufficiency

If we deny Strong Sufficiency, we have no explanation of why it is infelicitous to assert something that entails the falsity of $A \mathrel{\square\!\!\rightarrow} B$ but then also assert $A \Rightarrow B$.

2.3 Summarizing

We find ourselves in a pickle. The two bounding puzzles seem not easily avoided, at least without giving up something highly intuitive. Furthermore, it remains to be seen why indicative and subjunctive conditionals pattern so differently, with indicatives seeming to be stronger than the conditions under which they are believed, and subjunctives seeming to be weaker than the conditions under which they are believed. We need to find a new path.

Appendix

In Sarah Moss's theory (2015, 2018), sentences are assigned sets of credal measures m relative to a partition $\mathbb{Z}$ (a set of mutually exclusive propositions whose union is $\mathbf{W}$). m is a function that maps propositions to their probability according to that measure, where $m^{|\mathbf{A}} = m(\cdot|\mathbf{A})$. Moss defines the semantic value of a conditional $A \to B$ relative to such indices as follows (I ignore nested conditionals, so just assume A and B are both non-conditional; the presentation here follows Moss 2015):

$$[\![A \to B]\!]^{m,\mathbb{Z}} = 1 \text{ iff}$$
$$m(\bigcup\{\mathbf{P} \in \mathbb{Z} : m(\mathbf{B}|\mathbf{P}) = 1\} | \bigcup\{\mathbf{Q} \in \mathbb{Z} : m(\mathbf{A}|\mathbf{Q}) = 1\}) = 1$$

What this says is that a conditional $A \to B$ is true at $m, \mathbb{Z}$ iff the probability of the union of propositions in $\mathbb{Z}$ that probabilistically imply **B** (relative to m) conditional on the union of the propositions in $\mathbb{Z}$ that probabilistically imply **A** (relative to m) is 1.

Moss then defines a probability operator Pr relative to a measure function m and a partition $\mathbb{Z}$ that maps sets of measures $\mathcal{A}$ to values in the unit interval as follows:

$$Pr^m_{\mathbb{Z}}(\mathcal{A}) = m(\bigcup\{\mathbf{P} \in \mathbb{Z} : m^{|\mathbf{P}} \in \mathcal{A}\})$$

What this says is that the probability assigned to $\mathcal{A}$ equals the value m assigns to the union of propositions in $\mathbb{Z}$ such that conditionalizing

m on them yields a measure that's in $\mathcal{A}$. Note, then, that for Moss, the probability of a conditional $A \to B$ will always equal the probability of some proposition: namely, the union of propositions $\mathbf{P}$ in some partition such that the measure $m^{|\mathbf{P}}$ assigns 1 to the consequent conditional on its antecedent.

We'll see in Chapter 5 that Moss's theory nicely predicts the correct probabilities for certain conditionals in certain contexts (these are conditionals whose probabilities are not equal to the corresponding conditional probabilities of their consequents given their antecedents—see McGee 2000; Kaufmann 2004; Moss 2015). However, as we'll see in a moment, Moss's theory struggles to predict probabilities for conditionals that match Stalnaker's Thesis. Consider the following example: a ball is under exactly one of cups A, B, or C—that's all we know about the matter. Given this, our probability that the ball is under A is 1/3, and likewise for B and C. In this case, (24) is intuitively true, since the probability that the ball is under B given that it's not under A is 1/2:

(24) It's fifty-fifty likely that if the ball isn't under cup A, it's under cup B.

Can Moss's theory predict this? Whether it does will depend on what partition the *likely* operator is interpreted relative to. However, in this context, the only plausible partitions involve Boolean combinations of **A**, **B**, and **C** (that the ball is under A, etc). And the probability of any such Boolean combination is a multiple of 1/3. Of course, Moss could propose some partition $\{\mathbf{X}, \bar{\mathbf{X}}\}$ for the *likely* operator, where $m(\mathbf{X}) = 1/2$—but the issue here is whether there is any independent plausibility to such a move. After all, we might simply stipulate that, in this context, there is no proposition(s) around whose union has a probability is equal to 1/2. So, it seems, Moss's theory is unable to predict this intuition.

However, things get worse for Moss's theory. My choice of example here is inspired by Hájek 1989, 2011c, who shows how to generalize the result: for any finite domain of worlds, there will be more conditional probability values than unconditional probability values. That means that even if we can find a proposition whose probability is equal to some conditional probability in some cases, in a finite domain, there are guaranteed to be exceptions to Stalnaker's Thesis. This is surprising, especially since the kinds of contexts in which we do find intuitive exceptions to Stalnaker's Thesis are relatively rare, and even in those contexts, the non-Thesis-violating intuition is still accessible (more on this in Chapter 5).

For this reason, I think it is better to classify Moss's theory as one in which conditionals are predicted to have non-standard probabilities; Moss could then account for our intuitions when they diverge from these predictions by a kind of error theory—perhaps along the lines sketched in §2.1.2. I should also note that the restrictor gambit might be a particularly appealing strategy for Moss, since she predicts low-scope probabilified conditionals match Stalnaker's Thesis, in the following sense (similar to Kratzer 1986); see Moss 2015: 31–34:

If A, x-probably B is true (relative to m) iff $m(\mathbf{B}|\mathbf{A}) = x$.

We will return to Moss's theory when we discuss the exceptions to Stalnaker's Thesis in Chapter 5.

3
Domain Inferentialism

Chapters 1 and 2 articulated the bounding puzzles, and made the case for thinking that they are serious challenges facing any theory of conditionals. So far, so bad. Now, we turn to the good—in this chapter, I present my solution to the indicative and subjunctive bounding puzzles.

I argue for three claims in this chapter:

> <u>Claim 1</u>: Conditionals—both indicative and subjunctive—encode constraints on inferential dispositions.
>
> <u>Claim 2</u>: What it is to believe or leave open the possibility of a conditional is a matter of an agent's inferential dispositions, where these are entirely determined by her factual beliefs.
>
> <u>Claim 3</u>: The above claims provide an independent motivation for rejecting Conditonalization^{-} and Weak Sufficiency, while at the same time explaining what was compelling about them in the first place, thus avoiding the bounding puzzles.

Before we jump in, let me offer a brief preview of where we are going. The view I defend is **contextualist** (the content of a conditional depends on the contextual assignment of a modal base) and **non-factualist** (the content of a conditional will often carve finer-grained distinctions than those drawn by any possible fact), but **not expressivist**, at least in the sense of Yalcin and others (since the informational content of—that is, what is communicated by an utterance of—a conditional is a proposition). Here is Yalcin articulating the central tenet of modal expressivism:

> "In modeling the communicative impact of an epistemic possibility claim, we construe the objective as one of coordination on a certain

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0004

> global property of one's state of mind—the property of being compatible with a certain proposition—not one of coordination concerning the way the world is." (Yalcin 2011: 310)
>
> "In asserting something like
>
> (1) Allan is probably in his office,
>
> one may express an aspect of one's credal state, without describing that state. One expresses one's confidence, that is, without literally saying that one is confident. The relevant credal state expressed is of course a doxastic (hence 'cognitive') state, but it is not a state tantamount to full belief in a proposition. Credal expressivism is what I call this view." (Yalcin 2012a: 125)

Where expressivists see communication as coordinating on finer-grained contents (such as norms, plans, or credal states), my view is conservative—the information we use conditionals to coordinate is ordinary propositional information, even though conditionals very often have contents that are finer-grained, as we can see from considering truth-conditional entailments and the probabilities of conditionals.

3.1 Inferential Dispositions

There is a venerable idea, going back at least to Ramsey, that to believe a conditional is to believe its consequent conditional on believing its antecedent (cf. Ramsey 1931, Stalnaker 1984, Mellor 1993, Edgington 1995, Bennett 2003). This kind of view is typically associated with what I've called radical expressivist semantic theories, on which conditionals do not express propositional contents at all, and thus lack truth values and probabilities altogether (see the discussion in Chapter 2). Indeed, Mellor suggests as much when he says:

> " 'If P, Q' . . . expresses a disposition to infer Q from P. In other words, fully to accept a simple 'If P, Q' is to be disposed fully to believe Q if I fully believe P." (Mellor 1993: 236)

I think Mellor's insight is an important one, but I don't think we have yet appreciated how to correctly incorporate it into a theory of conditionals. Most clearly, this kind of approach to conditionals seems to be a version of radical expressivism, since there are no facts about whether to infer **B** from **A**, and thus such an approach is vulnerable to the arguments against radical expressivism discussed in Chapter 2.

My strategy aims to revive Mellor's insight, but adapt it in a different way. Like Mellor, I hold that conditionals encode constraints on inferential dispositions; however, I also think that they encode factual content (which for our purposes can be modeled as a possible worlds proposition). Thus, I think conditional contents are *mixed*—providing a constraint both on inferential dispositions and factual information.

Given this, we can characterize the conditions under which an agent accepts a conditional in terms of her inferential dispositions and factual beliefs. However, I will go a step further, and argue that an agent's inferential dispositions are entirely determined by her factual beliefs. Thus, we'll be able to state the acceptance conditions for conditionals entirely in terms of properties of her factual belief state. Furthermore, I'll argue that we can state every doxastic relation an agent might bear to a conditional (acceptance, rejection, leaving open the possibility of, and all the credal relations) in terms of properties of her factual beliefs. This chapter will focus entirely on non-graded doxastic relations like *belief* and *leaving open the possibility of*. I return to probabilistic relations in Chapter 5.

3.1.1 Conditionals and Inferential Dispositions

Let's define a view:

> **Simple Inferentialism**
> The contents of conditionals encode the disposition to infer their consequents from their antecedents.

Adherents of some (possibly restricted) version of this kind of view include theorists like Mellor (cited above), as well as Gibbard 1981, Stalnaker 1984.[1] However, while this view makes reasonable predictions for indicative conditionals, it stumbles when it comes to subjunctive conditionals. It is plausible that I believe (1) just if I believe that someone else shot Kennedy conditional on that Oswald didn't, and I have this conditional belief just because I believe someone shot Kennedy and leave open the possibility that someone other than Oswald did it.

(1) If Oswald didn't shoot Kennedy, someone else did.

However, this doesn't seem right for the subjunctive (2):

(2) If Oswald hadn't shot Kennedy, no one else would have.

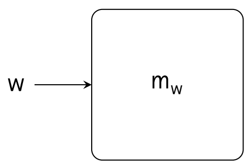
I may believe (2) even though I am not disposed to believe that no one else shot Kennedy conditional on that Oswald didn't. Indeed, this is clear from the fact that I may believe both (1) and (2) together.

For this reason, theorists drawn to conditional belief views about indicative conditionals tend to reject the analogous position for subjunctive conditionals, and thus adopt a disunified theory of conditionals (see, for instance, Gibbard 1981, Stalnaker 1984, Bennett 2003). But we have independent reason to think that indicative and subjunctive conditionals share a unified semantic core. For one, as we'll see, indicative and subjunctive conditionals share a common core logic, and their semantic differences seem to arise from the

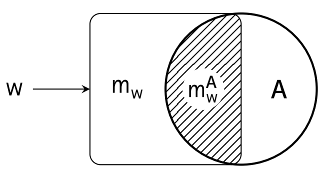
[1] Suppose that, given Simple Inferentialism, what it is to believe $A \to B$ just is to be disposed to believe B given belief in A; then, wouldn't any view that accepts Modus Ponens accept it? No, for two reasons. One is that even if Modus Ponens–validating theories accept this consequence of Simple Inferentialism, this doesn't mean that they accept Simple Inferentialism. A material conditional theory accepts this consequence but thinks that $A \to B$ encodes only the factual information either $\neg A$ or B. This brings us to the second point. The most natural view about the conditions under which one leaves open the possibility of a conditional encoding inferential dispositions is to not be disposed inconsistently with the dispositions it encodes. Thus, on Simple Inferentialism, to leave open the possibility of $A \to B$ just is to not be disposed to believe $\neg B$ given belief in A. But that is not the condition under which one leaves open the possibility of (for example) the material conditional $A \supset B$—one does the latter if one leaves open the possibility of $\neg A$ and this is compatible with being disposed to believe $\neg B$ given belief in A.

compositional contributions of their grammatical differences (see, for instance, Kratzer 1977, 1981; I take up this topic in more detail in Chapters 7–8). And second, adopting a unified theory on which both indicative and subjunctive conditionals encode inferential dispositions will allow for a unified solution to both bounding puzzles.

In light of these considerations, we should look for a more nuanced view about the inferential dispositions encoded by conditionals. To start, I propose that conditionals should be semantically analyzed as having a domain, partially determined by context, that is restricted by their antecedent. This will allow me to capture the semantic difference between indicative and subjunctive conditionals as a difference in their domains of quantification (see Chapter 8). Since the domain of a conditional is determined in part by its grammar and the context, the resulting view is **contextualist** (a point I will come back to in more detail later). For now, I will assume that the domain of a conditional is determined by a modal base m, which is a function from a world w to a set of worlds m_w (to be refined later, in Chapters 8–9), and not worry about how m is fixed by the context.



A conditional's domain is a set of antecedent-worlds compatible with the output of its domain function at some world. So, where A is the conditional's antecedent, its domain is: $m_w^{\mathbf{A}}$.



When features of the domain matter, I will write the conditional with its domain function subscripted, as in '$A \rightarrow_m B$'. In what follows, I will often use a subscripted conditional like this to denote an arbitrary conditional (both indicative and subjunctive).

To be refined below, the informal statement of my view is that $A \rightarrow_m B$ encodes the disposition to infer its consequent **B** from its domain together with its antecedent $m_w^{\mathbf{A}}$:

> **Domain Inferentialism**
> The content of a conditional encodes the disposition to infer its consequent from its domain together with its antecedent.

Unlike Simple Inferentialism, my version of inferentialism allows for a unified theory of the Oswald conditionals.

Suppose that the domain of the indicative conditional (1) at a world w is the set of worlds compatible with what is known (by a relevant group) at which Oswald didn't shoot Kennedy.

(1) If Oswald didn't shoot Kennedy, someone else did.

Then, according to my view, (1) encodes the disposition to infer that someone else shot Kennedy from that Oswald didn't and what's known continues to hold. If the knowledge of the relevant group is our own and we are introspective, to accept (1) just is to be disposed to infer that someone else shot Kennedy from that Oswald didn't.

Suppose that the domain of the subjunctive conditional (2) at world w is what's settled at w at some time t, just before Oswald shot Kennedy.

(2) If Oswald hadn't shot Kennedy, no one else would have.

Then the conditional encodes the disposition to infer that someone else shot Kennedy from the history of w until t and that Oswald didn't shoot Kennedy. To accept (2) just is to be disposed to infer that someone else shot Kennedy from the history of w until t and that Oswald didn't shoot Kennedy.

Someone might be disposed to infer that someone else shot Kennedy from their knowledge plus that Oswald didn't, but not be disposed to infer that someone else shot Kennedy from the history until t plus that Oswald didn't. Thus, we are in a position to explain how someone could believe (1) without believing (2) in a theory on which both encode inferential dispositions.

For this strategy to work, it is crucial that a conditional encodes the disposition to infer its consequent from its domain together with its antecedent. Thus, it is this version of domain inferentialism that I will be working with.

Summary: *Conditionals encode inferential dispositions*

It is plausible that indicative conditionals encode constraints on inferential dispositions; but this motivation extends to subjunctives too, since a unified theory of indicative and subjunctive conditionals is preferable, and this motivates Domain Inferentialism.

3.1.2 Belief and Inferential Dispositions

What is an inferential disposition? Let's start with an example. Suppose you have narrowed down the list of possible suspects to the butler and gardener; on that basis you are disposed to infer that the butler did it from that the gardener didn't, and vice versa. Knowing about this disposition of yours allows us to predict and make sense of your behavior. For instance, knowing you have this disposition, we'll expect you to believe the butler did it upon learning that you have come to believe the gardener didn't. This disposition also may figure into explanations of why you think evidence about the two suspects is linked in various ways.

Inferential dispositions are dispositions to believe whose triggering conditions are themselves beliefs: schematically, they may be described as dispositions to believe **B** upon believing **A**. But importantly, having the disposition to infer **B** from **A** is not the same thing as it being true of you that you would believe **B** were you to believe **A**, or that it would be rational for you to believe **B** were you to believe **A** (contra Levi 1978, Gärdenfors 1981). Bennett 2003: 28 helpfully distinguishes these when discussing the Ramsey Test, drawing on an example from van Fraassen 1980 (by way of Richmond Thomason).[2] Suppose that Jones has no belief one way or the other whether her business partner is cheating her. On that basis, Jones may accept:

(3) If my business partner is cheating me, I don't believe it.

On the kind of inferential dispositionalist theory I am building, to accept this is, roughly, for Jones to be disposed to infer that she doesn't believe her business partner is cheating her from that her business partner is cheating her. Jones may very well have this disposition, on account of her knowing that she doesn't believe one way or the other whether her business partner is cheating her. But notice that were Jones to believe that her business partner is cheating her, she would also believe that she believes this. Thus, Jones's disposition here is finkish (see Martin 1994, Lewis 1997)—coming to believe the triggering condition of the disposition removes the disposition entirely, since it removes its factual ground. The fink here is due to the fact that Jones is introspective: were she to come to believe that her business partner is cheating her, she would believe that she believes this. However, the fact that the disposition is finkish should not lead us to doubt that Jones does have the disposition. Indeed, it seems that she is so disposed merely because she knows that she doesn't believe her partner is cheating her—in other words, her disposition is entirely the result of her factual beliefs. Below, I will argue that this holds generally: the conditions under which one has an inferential disposition are entirely determined by their factual beliefs.

This point is critical for responding to a crucial worry about thinking that counterfactuals encode inferential dispositions. Stalnaker 1984 gives the following example. I accept that if Hitler had decided to invade England in 1940, Germany would have won the war. But if I

[2] Here is Ramsey's original formulation: "If two people are arguing 'If p will q?' and both are in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q. . . . We can say that they are fixing their degrees of belief in q given p" (Ramsey 1931: 249). I am not entirely sure what Ramsey meant by "adding p hypothetically" to one's stock of knowledge, but if he meant something that would trigger the disposition encoded by a conditional, then one may understand my view as a precisification and elaboration of the Ramsey Test.

were to learn that Hitler did in fact decide to invade England in 1940, I would not thereby believe that Germany did win the war. Rather, "my rejection of the antecedent was an essential presupposition to my acceptance of the counterfactual, and so gives me reason to give up the counterfactual rather than to accept its consequent, [were I to] learn the antecedent is true." (105–106).

This case bolsters my point that being disposed to infer that Germany won the war from that Hitler decided to invade England in 1940 is not the same as it being true of you that you would believe the former were you to learn the latter. In this case, your disposition to infer that Germany won from that Hitler decided to invade England is plausibly grounded in some factual belief of yours (perhaps about the composition of the German army compared to that of the English army in 1940), which, together with the claim that Hitler decided to invade, entails that Germany won.[3] But since your belief that Germany lost the war is more epistemically entrenched than your belief that Hitler didn't decide to invade England, learning the latter is false won't necessarily result in your revising the former—rather, in such a case, you would instead revise the factual belief that grounds the relevant inferential disposition.[4]

To distinguish the kind of contents that encode inferential dispositions from the kind of contents that encode merely factual information, I will call the former **refined contents** and the latter **proposi-**

[3] I defend the view that *all* of your inferential dispositions are grounded in some factual belief below.

[4] This example thus also illustrates the distinction between your inferential dispositions and how you would revise your beliefs upon learning something previously ruled out, which is one way of understanding the project of the belief revision literature: Gärdenfors 1988; Gärdenfors & Makinson 1988; Gärdenfors 2003; Hansson 2006, 2012 (see Blumberg & Lederman 2020 for this kind of interpretation of that literature). How you would revise your beliefs is sensitive to a measure of epistemic entrenchment encoding the information value of those beliefs for the agent—as Gärdenfors puts it, "When a belief set K is revised or contracted, the sentences in K that are given up are those having the lowest degrees of epistemic entrenchment" (Gärdenfors 2003: 17). But counterfactual inferential dispositions are not so sensitive. Suppose Jones is a committed atheist. We may also suppose that Jones has no factual beliefs that would ground a counterfactual inferential disposition to infer that the Christian God exists from that a god exists. As such, Jones does not believe the counterfactual *if there had been a god, it would have been the Christian God.* Yet, it may be true of Jones that were she to believe that a god exists, she would believe that the Christian God exists—perhaps because she was raised in a Christian household.

tional contents, and I will use uppercase teletype font for the former ($\mathtt{A}$, $\mathtt{B}$, $\mathtt{C}, \ldots$) and uppercase boldfaced font for the latter (**A**, **B**, **C**, ...). Although these are different in kind, they bear an important relationship to one another. Propositional contents can be identified with refined contents that encode only trivial inferential dispositions, while refined contents determine propositional contents (more on this below). When two refined contents differ in their inferential dispositions but not their propositional content, they differ non-factually.

Conditionals may be non-factual, and thus we need refined contents to model their meanings. But, then, what is it to believe or leave open the possibility of a refined content? A natural idea is that you believe a refined content when you are inferentially disposed in accordance with it. Consider for a moment the simple (and incorrect, as we saw above) view that the indicative conditional $A \rightarrow B$ encodes the disposition to infer **B** from **A**. According to that view, to believe $A \rightarrow B$ just is to be disposed to infer **B** from **A**. Suppose further that it's sufficient for being disposed to infer **B** from **A** that you believe the material conditional $\mathbf{A} \supset \mathbf{B}$ and do not believe $\bar{\mathbf{A}}$ (this is something I'll argue for below). Then, it follows that it is sufficient for believing $A \rightarrow B$ that you believe the material conditional $\mathbf{A} \supset \mathbf{B}$ and do not believe $\bar{\mathbf{A}}$. Furthermore, when you don't believe $\bar{\mathbf{A}}$, coming to believe $A \rightarrow B$ would require learning nothing more than $\mathbf{A} \supset \mathbf{B}$.

These are very nice results, predicting both Weak Sufficiency and Stalnaker 1975's Or to If principle, although they do not follow without the further step of reducing inferential dispositions to factual beliefs—something I take up in the next section.

Weak Sufficiency

Coming to accept $A \rightarrow B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.

Or to If

An agent who accepts $A \vee B$ without accepting either disjunct thereby accepts $\neg A \rightarrow B$.

Now, we come to the complications. First, remember, I endorse Domain Inferentialism, so the dispositions encoded by $A \rightarrow_m B$ are to infer **B** from $m_w^{\mathbf{A}}$. But also, remember that conditional domains vary across worlds, and thus, "the domain of $A \rightarrow_m B$" is really a random variable—a function from worlds to propositions. So, strictly speaking, a conditional encodes a disposition to infer its consequent from a random variable. What this means is that the refined content of a conditional encodes both factual information and inferential dispositions simultaneously:

Domain Inferentialism (Precise)
$A \rightarrow_m B$ encodes, for each w, the disposition to infer **B** from $m_w^{\mathbf{A}}$.

This is an important clarification, because it does seem that conditionals don't *merely* encode inferential dispositions. Here is an example. Of a never flipped coin, consider:

(4) If it had been flipped, it would have landed heads.

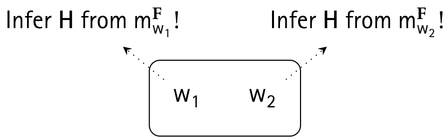
$$F \Rightarrow H$$

The dispositions encoded by $F \Rightarrow H$ vary depending on the world. At w_1, where the coin is double-headed, $F \Rightarrow H$ encodes the trivial disposition to infer **H** from $m_{w_1}^{\mathbf{F}}$. This disposition is trivial because its domain $m_{w_1}^{\mathbf{F}}$ entails both that the coin is double-headed and the law that double-headed coins land heads if flipped (we are supposing), and thus $m_{w_1}^{\mathbf{F}} \subseteq \mathbf{H}$. Thus, this disposition is trivial—everyone has it trivially. At w_2, where the coin is fair, $F \Rightarrow H$ encodes the non-trivial disposition to infer **H** from $m_{w_2}^{\mathbf{F}}$. This disposition is non-trivial because $m_{w_2}^{\mathbf{F}} \nsubseteq \mathbf{H}$ and $m_{w_2}^{\mathbf{F}} \nsubseteq \bar{\mathbf{H}}$. So, not every individual will have this disposition (in fact, I'll argue below that no individual has this disposition!).

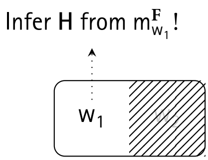
Given this, and the guiding insight from earlier—that you believe a refined content if you are inferentially disposed in accordance with it—whether you're disposed in accordance with $F \Rightarrow_m H$ will depend both on your inferential dispositions and your factual beliefs. That is, if w_1 is compatible with your beliefs, then you will need to be disposed to infer **H** from $m_{w_1}^{\mathbf{F}}$; and if w_2 is compatible with your beliefs, you will need to be disposed to infer **H** from $m_{w_2}^{\mathbf{F}}$. Thus, if both worlds are compatible with your factual beliefs, you'll either have to have both inferential dispositions, or eliminate one or the other worlds and have whatever inferential dispositions from the conditional remain. Where $\mathbf{BEL}_S$ is S's belief state (the set of worlds compatible with what S believes):

> **Conditional Belief**
> S believes a conditional $A \rightarrow_m B$ iff for each $w \in \mathbf{BEL}_S$: S is disposed to infer **B** from $m_w^{\mathbf{A}}$.

In the case above, suppose our agent S starts out with the following factual belief state (with the corresponding dispositions encoded by the conditional at each world marked accordingly):



Let's suppose that our agent has the first disposition (the trivial one, to infer **B** from $m_{w_1}^{\mathbf{F}}$) but not the second. Then, to believe the conditional, he must either eliminate w_2 from his belief state, or come to have the second disposition. Suppose he does the former—that is, he comes to believe that the coin is double-headed; then his belief state will look like this and he will believe the conditional $F \Rightarrow_m H$:



Similarly, then, to not be inferentially disposed against the dispositions encoded by a conditional is for there to be a world compatible with your beliefs where the disposition encoded by the conditional relative to that world is compatible with your inferential dispositions:

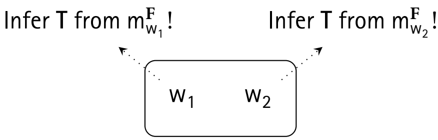
Conditional Possibility
S leaves open the possibility of $A \rightarrow_m B$ iff for some $w \in \mathbf{BEL}_S$: S is not disposed to infer $\bar{\mathbf{B}}$ from $m_w^{\mathbf{A}}$.

Now consider the conditional about the same coin:

(5) If it had been flipped, it would have landed tails.

$$F \Rightarrow_m T$$

Intuitively, S leaves open the possibility of this conditional, since he leaves open the possibility that the coin is fair. This is borne out by Conditional Possibility. To see why, consider S's initial belief state, where he leaves open the possibility that the coin is double-headed and that the coin is fair (here, I now indicate the dispositions encoded by the conditional $F \Rightarrow_m T$):



Now, remember that S trivially has the disposition to infer **H** from $m_{w_1}^{\mathbf{F}}$, since the domain at w_1 entails **H**; hence, assuming heads and tails are the only possible outcomes, this cannot be a world that witnesses the condition of Conditional Possibility. However, we may suppose that S lacks both the disposition to infer **H** from $m_{w_2}^{\mathbf{F}}$ and to infer **T** from $m_{w_2}^{\mathbf{F}}$. If that's so, then the fact that w_2 is compatible with S's beliefs is sufficient for S to leave open the possibility of $F \Rightarrow_m T$. And this seems exactly right!

Summary: *Believing conditionals*

What it is to believe/leave open the possibility of a conditional is a matter of the agent's inferential dispositions. We saw how one can bring one's inferential dispositions in line with a conditional (thus believing it) merely by having certain factual beliefs.

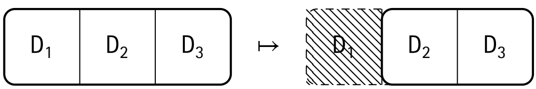
3.1.3 Grounding Inferential Dispositions

There is one final step in our theory, which is to say where an agent's inferential dispositions come from. On this score, I will argue that one's factual beliefs fully determine one's inferential dispositions. It is this step that will allow us to reduce belief in refined contents to a property of one's factual belief state.

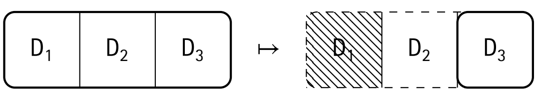
I argue for this claim by arguing that, as a matter of what inferential dispositions are, an agent's inferential dispositions must at least distinguish possibilities compatible with her factual beliefs from possibilities her factual beliefs rule out, must not distinguish within the possibilities compatible with her factual beliefs, and must not distinguish within the possibilities ruled out by her factual beliefs. I'll elaborate and motivate each point in turn.

1. If **A** is compatible with your factual belief state **S**, then you are disposed to infer **B** from **A** iff **A** ⊃ **B** is entailed by **S**.

    Suppose you think the prize is either behind door 1, 2, or 3, and that is all you think about the matter. Then, you are disposed to infer that the prize is behind doors 2 or 3 from that it isn't behind door 1 (and so on for 2 from not 1 or 3, etc.).

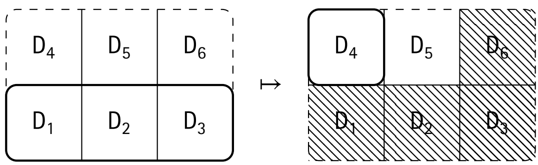
    Similarly, you are not disposed to infer that the prize is behind door 3 from that it isn't behind door 1 (doing so neglects a live possibility by your lights—that it's behind door 2).

These examples motivate the first two claims: your inferential dispositions must distinguish worlds compatible with your beliefs from those incompatible with your beliefs. You have the first disposition (to infer $\mathbf{D_2} \vee \mathbf{D_3}$ from $\bar{\mathbf{D}}_{\mathbf{1}}$) because among the $\bar{\mathbf{D}}_{\mathbf{1}}$-worlds in your belief state, all of them are $\mathbf{D_2} \vee \mathbf{D_3}$-worlds, some of them are $\mathbf{D_2}$-worlds, and some are $\mathbf{D_3}$-worlds. You lack the second disposition (to infer $\mathbf{D_3}$ from $\bar{\mathbf{D}}_{\mathbf{1}}$) because it would distinguish within possibilities that are each live by your lights (namely, $\mathbf{D_2}$ and $\mathbf{D_3}$-worlds).

2. If **A** is not compatible with your factual belief state **S**, then you are not disposed to infer **B** from **A**, unless $\mathbf{A} \subseteq \mathbf{B}$ (in which case you will have that disposition trivially).

Suppose you think the prize is not behind doors 4, 5, or 6. Then, you are not disposed to infer that the prize is behind door 4 from that it's behind door 4 or 5. But you are disposed to infer that the prize is behind door 4 or 5 from that it's behind door 4 (since $\mathbf{D_4} \subseteq \mathbf{D_4} \vee \mathbf{D_5}$).



This example motivates the third claim: your inferential dispositions do not distinguish between worlds that you've ruled out. You lack the disposition to infer $\mathbf{D_4}$ from $\overline{\mathbf{D_1} \vee \mathbf{D_2} \vee \mathbf{D_3} \vee \mathbf{D_6}}$) because it distinguishes between possibilities ruled out by your lights (namely, $\mathbf{D_4}$ and $\mathbf{D_5}$-worlds).

We thus establish:

Factual Determination
An agent S is disposed to infer **B** from **A** iff

a. **A** is compatible with S's factual beliefs and $\mathbf{A} \supset \mathbf{B}$ is entailed by them, or
b. **A** is incompatible with S's factual beliefs and $\mathbf{A} \subseteq \mathbf{B}$.

Notice that, again, to be disposed to infer **B** from **A**, when you've ruled out **A**, is not the same as it being true of you that you would believe **B** were you to learn **A**—recall the case of Jones the committed atheist, who has no dispositions to infer that the Christian God exists from that a god exists but nonetheless would have believed that the Christian God existed were she to have believed that a god existed, since she was raised in a Christian household.

We are now in a position to reduce our notions of believing or leaving open a conditional to properties of an agent's factual belief state. From Conditional Belief and Factual Determination, we have:

> **Conditional Belief (Reduced)**
> S believes $A \rightarrow_m B$ iff for each world w compatible with S's factual beliefs,
>
> a. $m_w^{\mathbf{A}}$ is compatible with S's factual beliefs and $m_w^{\mathbf{A}} \supset \mathbf{B}$ is entailed by them, or
> b. $m_w^{\mathbf{A}}$ is incompatible with S's factual beliefs and $m_w^{\mathbf{A}} \subseteq \mathbf{B}$.

And from Factual Determination and the fact that leaving open is the dual of belief, we have:

> **Leaving Open (Reduced)**
> S leaves open the possibility of $A \rightarrow_m B$ iff for some world w compatible with S's factual beliefs,
>
> a. if $m_w^{\mathbf{A}}$ is compatible with S's factual beliefs, then $m_w^{\mathbf{A}} \supset \bar{\mathbf{B}}$ is not entailed by them, and
> b. if $m_w^{\mathbf{A}}$ is incompatible with S's factual beliefs, then $m_w^{\mathbf{A}} \not\subseteq \bar{\mathbf{B}}$.

Suppose you know that Smith had a fair coin, but don't know whether Smith flipped it or, if he did, how it landed.

(6) If Smith flipped the coin, it landed heads.

$$F \rightarrow_m H$$

Again, suppose you know the conditional's domain is $m_w^{\mathbf{F}}$—the set of flip-worlds compatible with your knowledge. You don't believe the conditional because $m_w^{\mathbf{F}}$ is compatible with what you believe and

$m_w^{\mathbf{F}} \supset \mathbf{H}$ is not entailed by your beliefs. But you do regard it an open possibility because $m_w^{\mathbf{F}}$ is compatible with what you believe and $m_w^{\mathbf{F}} \supset \bar{\mathbf{H}}$ is not entailed by your beliefs.

Now suppose you learn that Smith didn't flip the coin, and consider the subjunctive conditional:

(7) If Smith had flipped the coin, it would have landed heads.

$$F \Rightarrow_m H$$

In this case, the conditional's domain varies across worlds compatible with your knowledge. At a double-heads world w_1, its domain is $m_{w_1}^{\mathbf{F}} \subseteq \mathbf{H}$; and at a double-tails world w_2, its domain is $m_{w_2}^{\mathbf{F}} \subseteq \mathbf{T}$. You don't believe this conditional because $m_{w_2}^{\mathbf{F}}$ is not compatible with what you believe and $m_{w_2}^{\mathbf{F}} \nsubseteq \mathbf{H}$. And you leave open the possibility of this conditional because $m_{w_1}^{\mathbf{F}}$ is incompatible with what you believe and $m_{w_1}^{\mathbf{F}} \nsubseteq \bar{\mathbf{H}}$.

Notice, finally, that we have stated what it is to believe or leave open a conditional entirely in terms of properties of a factual belief state, thus reducing the doxastic properties of refined contents to doxastic properties of propositions. And we generated this result by (i) motivating the view that belief in or leaving open a refined content was a feature of an agent's inferential dispositions, and (ii) motivating the view that an agent's inferential dispositions are determined by their factual beliefs.

Summary: *Grounding inferential dispositions*

An agent's inferential dispositions are grounded in her factual beliefs. Given this, we can state what it is to believe or leave open the possibility of a conditional in terms of an agent's factual beliefs.

3.1.4 Constitutivism vs. Rationalism

Before we see how this theory handles non-factual ignorance and our bounding puzzles, I want to return briefly to the Or to If inference pattern in order to articulate a contrasting strategy with the one I have pursued here, and offer some motivation for my preferred approach.

> **Or to If**
>
> An agent who accepts $A \vee B$ without accepting either disjunct thereby accepts $\neg A \rightarrow B$.

Some evidence for Or to If comes from the fact that conjunctions like the following are infelicitous:

(8) #Smith is either in Athens or in Barcelona; and if he's not in Athens, he's not in Barcelona either.

Above, I motivated Weak Sufficiency along similar grounds. Pushing back on this, someone might propose an alternative explanation of why (8) sounds odd—it might be unassertable because the state of mind it expresses is irrational. Generalizing this strategy, we might hold instead that Weak Sufficiency and Or to If are requirements of rational belief states. Thus, distinguish two strategies (cf. Street 2008's discussion of what constitutes taking oneself to have a reason to ϕ).

- According to Strategy 1, to believe $A \rightarrow B$ is to be inferentially disposed in accordance with it; to believe $A \supset B$ while leaving open the possibility of A is sufficient to be inferentially disposed in accordance with $A \rightarrow B$; so violations of Or to If are impossible.
- According to Strategy 2, to believe $A \rightarrow B$ is to be inferentially disposed in accordance with it; but to believe $A \supset B$ while leaving open the possibility of A is sufficient only for it to be rational for you to be inferentially disposed in accordance with $A \rightarrow B$. So, violations of Or to If are possible, but irrational.

Strategy 1 is the approach I am taking in this book. But I want to point out that almost everything I say can be reformulated in Strategy 2 terms if that is what the evidence calls for. The main reason I have for preferring Strategy 1 is that the infelicity of (8) extends to embeddings in contexts in which it is unasserted:

(9) a. #Jones thinks that Smith is either in Athens or in Barcelona and that if he's not in Athens, he's not in Barcelona either.
b. #Suppose that Smith is either in Athens or in Barcelona and that if he's not in Athens, he's not in Barcelona either.

Here, no irrational state of mind is expressed; rather, such a state is attributed to Jones and commanded to be supposed. But there should be nothing wrong with doing that; there is nothing wrong with attributing irrational beliefs to someone, or supposing them:

(10) a. Jones thinks that Dasher will win the race and that he has no evidence that Dasher will win.
b. Suppose that Dasher will win the race and that you have no evidence that Dasher will win.

The infelicity of (8), (9-a), (9-b) thus suggests, I think, that it simply is not possible to believe $A \vee B$ (while leaving open $\neg A$) without believing $\neg A \rightarrow B$. On this hypothesis, the oddity of these sentences is that they express, attribute, and suppose a content that cannot be believed or supposed.[5]

3.2 Leaving Open the Possibility of Conditionals

Consider the following case.

> **Draw or Flip**. Joe and Sue face a choice between flipping a coin and drawing the top card of a freshly shuffled deck (suppose each has their own deck). Each knows that if they draw a red card, they'll get \$1, and if they draw black, they'll get \$0. Each also knows that if they flip heads, they'll get \$1, and if they flip tails, they'll get \$0.

Suppose that Sue flips and gets tails; now consider:

[5] Of course, someone might be confused and think they believe (8). But merely sincerely asserting it is not sufficient reason to think that they really do believe it.

(11) a. If Sue had drawn, she would have got red. $D \Rightarrow R$
b. If Sue had drawn, she would not have got red.

And suppose Joe draws the top card and gets black; now consider:

(12) a. If Joe had flipped, he would have got heads. $F \Rightarrow H$
b. If Joe had flipped, he would not have got heads.

Three observations. First, we don't know either conditional, and thus we leave open the possibility of each of (11-a)/(11-b) and (12-a)/(12-b). And the fact that we leave open the possibility of both of (11-a)/(11-b) is sufficient for us to leave open the possibility that Sue's choice to flip is regrettable and that her choice to flip is not regrettable—whether it is depends on the color of the top card of her deck. If the top card was red, then her choice to flip was regrettable, but if the top card was black, then her choice was not regrettable. By contrast, the fact that we leave open the possibility of both of (12-a)/(12-b) is compatible with us fully believing, and even knowing, that Joe's choice to draw was regrettable. Since the top card of his deck was black, he was facing a choice between guaranteed nothing and a 0.5 chance at $1 and a 0.5 chance at nothing.

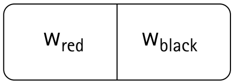
Second, our ignorance about (11) is resolvable: once we learn the color of the top card of Sue's deck, we learn which of these conditionals was true. By contrast, our ignorance of (12) is unresolvable: since we know the coin was fair and unflipped, there was nothing we can learn to come to believe which of these conditionals is true and which is false.

Third, these first two observations seem clearly related. Our ignorance about which of (11) is true is ignorance about some fact, which is why it is sufficient for ignorance about the regrettability of Sue's choice. But our ignorance about which of (12) is true is not ignorance about some fact (but rather, merely a reflection of our inferential dispositions), which is why it is not sufficient for ignorance about the regrettability of Joe's choice.[6]

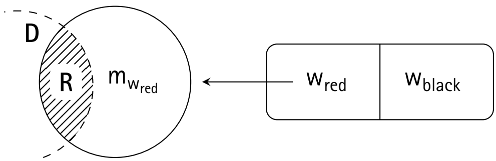
In this section, I will draw on the domain inferentialist theory sketched in the previous sections to explain this pattern of observa-

[6] For a more detailed version of this argument, see Khoo 2021.

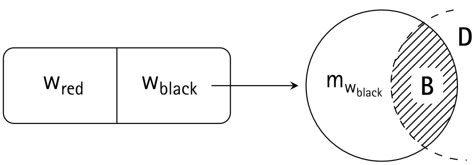
tions. Start with (11). We don't know what the top card of Sue's deck is, but it is either red or black, so we can partition the worlds in your belief state into two cells: w_{red} where the top card was red and w_{black} where the top card was black:



Drawing the top card when it is red ensures that Sue will draw red. Therefore, at w_{red}, we have: $m^{\mathbf{D}}_{w_{red}} \subseteq \mathbf{R}$. And thus, you are (trivially) disposed to infer $\mathbf{R}$ from $m^{\mathbf{D}}_{w_{red}}$.



Drawing the top card when it is black ensures that Sue will draw black. Therefore, at w_{black}, we have: $m^{\mathbf{D}}_{w_{black}} \subseteq \bar{\mathbf{R}}$. And thus, you are (trivially) disposed to infer $\bar{\mathbf{R}}$ from $m^{\mathbf{D}}_{w_{black}}$.



As a result, you neither believe $D \Rightarrow R$ nor $D \Rightarrow \neg R$, and instead you leave open the possibility of each. I'll focus on the former.

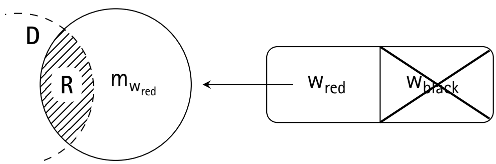
Why don't you believe $D \Rightarrow R$?
To believe this conditional is to be disposed to infer $\mathbf{R}$ from $m^{\mathbf{D}}_{w}$, for each w compatible with your beliefs. But w_{black} is compatible with your beliefs, and you are not disposed to infer $\mathbf{R}$ from $m^{\mathbf{D}}_{w_{black}}$ (that Sue drew a red card from that Sue drew and the top card was black).

Why do you leave open the possibility of $D \Rightarrow R$?
To leave open this conditional is for there to be worlds w compatible with your beliefs such that you're not disposed to infer $\bar{\mathbf{R}}$ from $m^{\mathbf{D}}_{w}$.

And there is a world compatible with your beliefs, namely, w_{red}, at which you are not disposed to infer $\bar{\mathbf{R}}$ from $m^{\mathbf{D}}_{red}$. This is so because you are in fact disposed to infer **R** upon believing that Sue drew and the top card was red.

This is the easy case, since it is a case of factual ignorance, meaning there is a fact of the matter about which conditional is true and thus you could come to believe the conditional to be true or to be false.

What would it take to believe $D \Rightarrow R$?
To believe this conditional is for your factual beliefs to be compatible with worlds w at which you are disposed to infer **R** from $m^{\mathbf{D}}_{w}$. As we saw above, you are not disposed to infer **R** from $m^{\mathbf{D}}_{w_{black}}$, but you are disposed to infer **R** from $m^{\mathbf{D}}_{w_{red}}$. So, if you rule out w_{black}, then you would be disposed to infer **R** from $m^{\mathbf{D}}_{w}$, for each w in your belief state. So, ruling out w_{black} is necessary and sufficient for believing $D \Rightarrow R$.

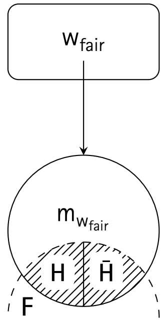


Turn next to (12), which I claim you are non-factually ignorant of:

(12) If Joe had flipped, he would have got heads.

$F \Rightarrow H$

Your know the coin is fair, so your belief state contains only one relevant type of world: w_{fair}, where the coin is fair. Since flipping a fair coin does not determine an even outcome or an odd outcome, we have that $m^{\mathbf{F}}_{w_{fair}} \not\subseteq \mathbf{H}$ and $m^{\mathbf{F}}_{w_{fair}} \not\subseteq \bar{\mathbf{H}}$. Since **F** is incompatible with your beliefs (you know Joe didn't flip), you are neither disposed to infer **H** from $m^{\mathbf{F}}_{w_{fair}}$ nor $\bar{\mathbf{H}}$ from $m^{\mathbf{F}}_{w_{fair}}$.



As a result, you neither believe $F \Rightarrow H$ nor $F \Rightarrow \neg H$; instead, you leave open the possibility of each. Again, I'll focus on the former.

Why don't you believe $F \Rightarrow H$**?**
To believe this conditional is for your belief state to only allow worlds w such that you are disposed to infer **H** from $m_w^{\mathbf{F}}$. But you are not disposed to infer **H** from $m_{fair}^{\mathbf{F}}$.
Why do you leave open $F \Rightarrow H$**?**
To leave open this conditional is for your factual beliefs to be compatible with worlds w such that you are not disposed to infer $\bar{\mathbf{H}}$ from $m_w^{\mathbf{F}}$. And you are not disposed to infer $\bar{\mathbf{H}}$ from $m_{fair}^{\mathbf{F}}$.

Finally, we can now show that this is indeed a case in which you leave open the possibility of $F \Rightarrow H$ but cannot come to believe it or its negation.

What would it take to believe $F \Rightarrow H$**?**
To believe this conditional is for your belief state to only allow worlds w such that you are disposed to infer **H** from $m_w^{\mathbf{F}}$. And since **F** is ruled out, the only way to do this would be to rule out the world w_{fair} where you lack the disposition to infer **H** from $m_{w_{fair}}^{\mathbf{F}}$. But this would trivialize your belief state. Therefore, you cannot come to believe this conditional, since there is nothing you can learn to ensure your inferential dispositions will match those it encodes.

Summary: *Leaving open conditionals*

Domain inferentialism predicts that there are conditionals such that our leaving them open is a matter of our factual ignorance; it also predicts that there are conditionals such that our leaving them open is not a matter of our factual ignorance. We predict that the former are believable but the latter not.

3.3 Resolving the Bounding Puzzles

3.3.1 Subjunctive Bounding Resolved

Recall the puzzle:

Strong Sufficiency

If you are sure that A is false, then you accept $A \Rightarrow B$ only if you believe $A \mathbin{\Box\!\!\rightarrow} B$.

Weakness

It is possible that $A \Rightarrow B$ is true and $A \mathbin{\Box\!\!\rightarrow} B$ is false.

Conditionalization$^-$

To come to believe A is at most to rule out all possibilities in which A is false.

From Weakness, there are possibilities in which the subjunctive $A \Rightarrow B$ is true and its strong subjunctive counterpart $A \mathbin{\Box\!\!\rightarrow} B$ is false (remember, the latter is true iff all of the worlds in $m_w^{\mathbf{A}}$ are **B**-worlds). So, there may be an agent S who does not rule out some such possibilities. By Conditionalization$^-$, if S were to come to believe $A \Rightarrow B$, she must not rule out those possibilities, but that would mean she would believe $A \Rightarrow B$ without believing $A \mathbin{\Box\!\!\rightarrow} B$, contradicting Strong Sufficiency.

My theory predicts Strong Sufficiency and Weakness, and invalidates Conditionalization$^-$. The case where Joe didn't flip the fair coin can be used to illustrate all three results:

(13) If Joe had flipped, he would have got heads. $F \Rightarrow H$

In this case, you leave open the possibility of $F \Rightarrow H$ even though you rule out its strong subjunctive counterpart $F \boxright H$ (since you know the coin is fair), thus illustrating Weakness. But furthermore, you cannot learn $F \Rightarrow H$ without also learning $F \boxright H$ (which you cannot learn at all since you know the coin is fair), thus illustrating Strong Sufficiency and thus invalidating Conditionalization$^-$. Here are the general proofs:

Predicting Strong Sufficiency
Remember that **F** is incompatible with your beliefs. Then, by Conditional Belief (Reduced), you believe $F \Rightarrow H$ only if for each w in your belief state, $m_w^{\mathbf{F}} \subseteq \mathbf{H}$. But that is sufficient for believing the strong subjunctive $F \boxright H$.

Predicting Weakness
It is possible for an agent to leave open the possibility of $F \Rightarrow H$ and rule out the strong subjunctive $F \boxright H$. By Leaving Open (Reduced), you leave open the possibility of $F \Rightarrow H$ if there is a world compatible with your beliefs w at which $m_w^{\mathbf{F}} \not\subseteq \bar{\mathbf{H}}$. And this is compatible with ruling out the strong subjunctive $F \boxright H$—to do so just is to rule out every world compatible with your beliefs w is such that $m_w^{\mathbf{F}} \not\subseteq \mathbf{H}$.

Invalidating Conditionalization$^-$
This follows from Weakness and Strong Sufficiency.

The key takeaway from all this is not merely that we have a way to falsify one of the premises of an inconsistent triad, but that we *independently motivated* our rejection of Conditionalization$^-$ by grounding belief in a conditional in an agent's inferential dispositions and grounding an agent's inferential dispositions in her factual beliefs. Certain ways of being inferentially disposed require believing some factual content that is sufficient to ground certain other inferential dispositions, thus ensuring that we end up believing something stronger than we would if we could just directly believe the conditional by minimally adjusting our inferential dispositions without meddling with our factual beliefs.

3.3.2 Indicative Bounding Resolved

We turn finally to the indicative bounding puzzle. Recall that the puzzle arises from three principles:

> **Weak Sufficiency**
>
> Coming to accept $A \to B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.
>
> **Strength**
>
> It is possible that $A \supset B$ is true and $A \to B$ is false.
>
> **Conditionalization^{+}**
>
> Coming to believe A is at least to rule out all possibilities in which A is false.

There are two steps to resolving the indicative bounding puzzle. The first is to make explicit a constraint on Weak Sufficiency that was implicit in our limiting it to indicative conditionals. Remember that, according to my theory, indicative and subjunctive conditionals are distinguished semantically by their domains. I will say more about that difference in Chapters 7–9, but here I want to articulate the crucial difference that ensures Weak Sufficiency has a chance of being true.

To start, notice that, without some constraint on the range of possible modal bases for indicative conditionals, counterexamples to Weak Sufficiency will be easy to come by. Let w be a $\bar{\mathbf{A}}$-world such that $m_w^{\mathbf{A}} \subseteq \bar{\mathbf{B}}$. Suppose S allows such possibilities in her belief state. Then, coming to believe $A \to_m B$ will require ruling out such possibilities (since S is trivially disposed to infer $\bar{\mathbf{B}}$ from $m_w^{\mathbf{A}}$, and this disposition is incompatible with the one encoded by the conditional). But, then coming to believe $A \to_m B$ will require believing more than just that $A \supset B$ is true, contra Weak Sufficiency.

This does not mean that we have solved the indicative bounding puzzle, for we have no explanation of why Weak Sufficiency is plausible. Rather, I will hold that a weakened version of Weak Sufficiency is true, on which the bounding puzzle still arises—and show that the domain inferentialist theory nonetheless can avoid it.

We want Weak Sufficiency for indicatives but not subjunctive conditionals, so, given that the semantic difference between them is due to differences in their modal bases, our strategy will be to weaken Weak Sufficiency to hold only for conditionals relative to certain kinds of modal bases.

Intuitively, indicative conditionals concern what is epistemically possible for an agent, which is typically presupposed to be lower bounded by that agent's factual beliefs. This means that an agent will typically only consider interpreting an (epistemic) indicative conditional relative to a modal base that she believes is no stronger than her beliefs—that is, that does not rule out any worlds compatible with her beliefs:

$$m \text{ is bounded by } \mathbf{BEL} \text{ iff } \forall w \in \mathbf{BEL} : \mathbf{BEL} \subseteq m_w.$$

Hence, if your belief state is $\{w_1, w_2\}$, the only candidates for epistemic modal bases for indicative conditionals will be those m such that m_{w_1} and m_{w_2} contain at least w_1, w_2. The reason for this is that we typically care about epistemic possibilities as a guide to what to believe—epistemic possibilities represent those compatible with some evidence that we are using to orient our beliefs. Therefore, if we think a modal base rules out some possibilities incompatible with our beliefs, we will either think that it represents "evidence" that is getting things wrong (ruling out live possibilities) or that it is accurate, in which case we should update our beliefs accordingly.[7]

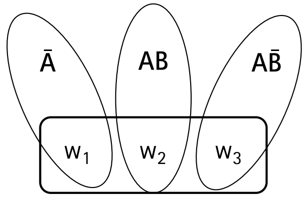
We can now state Weak Sufficiency more carefully as follows:

Weak Sufficiency*

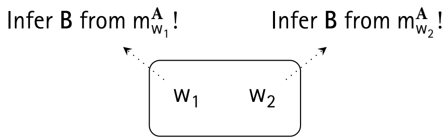
Coming to accept $A \rightarrow_m B$ (when you do not rule out **AB**) and m is bounded by your beliefs is to come to believe nothing more than that $A \supset B$ is true.

[7] I do leave it open as possible to interpret an epistemic modal expression (even an indicative conditional, perhaps) relative to an m that is not bounded by your beliefs. This might happen if we had some independent reason to care about the properties of that m, even recognizing that it may not be accurate. But I think such cases are a minority (see, for instance, Egan et al. 2005).

So, that's the first, clarificatory, point. Notice that by doing this, we immediately avoid the simple counterexample to Weak Sufficiency above. Suppose S leaves open the possibility of $\mathbf{AB}$, $\mathbf{A\bar{B}}$, and $\mathbf{\bar{A}}$:



Then any m bounded by S's beliefs will include at least these worlds (relative to w_1, w_2, w_3). And thus, there will be no worlds w compatible with S's beliefs such that $m_w^{\mathbf{A}} \subseteq \mathbf{\bar{B}}$. So, the minimal factual information S needs to learn to come to believe $A \rightarrow_m B$ (for any bounded m) is $\mathbf{A} \supset \mathbf{B}$—that is, S only need rule out w_3. Why? Consider the inferential dispositions encoded by $A \rightarrow_m B$ at w_1, w_2:



By ruling out w_3, you thus have both inferential dispositions. The reasoning is the same for both, so let's just look at the disposition to infer $\mathbf{B}$ from $m_{w_1}^{\mathbf{A}}$ (the conditional's domain at w_1). Since m is bounded by S's belief state, which is compatible with $\mathbf{A}$, $m_{w_1}^{\mathbf{A}}$ must contain at least w_2. Thus, the conditional's domain at w_1 is compatible with S's beliefs, and so, by Factual Determination (repeated below), S is disposed to infer $\mathbf{B}$ from the conditional's domain at w_1 if her belief state, together with the conditional's domain at w_1, entails $\mathbf{B}$. And it does, since the intersection of her belief state together with the conditional's domain is $\mathbf{BEL}_S \cap m_{w_1}^{\mathbf{A}} = \{w_2\}$, which entails $\mathbf{B}$.

Factual Determination
An agent S is disposed to infer $\mathbf{B}$ from $\mathbf{A}$ iff

a. $\mathbf{A}$ is compatible with S's factual beliefs and $\mathbf{A} \supset \mathbf{B}$ is entailed by them, or
b. $\mathbf{A}$ is incompatible with S's factual beliefs and $\mathbf{A} \subseteq \mathbf{B}$.

Thus, our theory falsifies Weak Sufficiency, but predicts Weak Sufficiency*. However, merely restricting the principle in this way is not how we avoid the indicative bounding puzzle. Notice that the reasoning that seemed to establish the inconsistency between the principles (from Chapter 1) carries over with Weak Sufficiency*:

> From Strength, there are possibilities in which $A \supset B$ is true and $A \rightarrow_m B$ is false. Then, there may be an agent S who does not rule out some such possibilities. Then, by Conditionalization$^+$, if S were to come to believe $A \rightarrow_m B$, she must rule out those possibilities, but that would be to rule out some possibilities in which $A \supset B$ is true, and hence come to believe something more than that $A \supset B$ is true, contradicting Weak Sufficiency and Weak Sufficiency*.

So, how does the theory avoid the puzzle? The issue here is how we interpret "possibility." The strength of indicative conditionals is measured by the fact that one can leave open the possibility of $A \supset B$ while ruling out the possibility of $A \rightarrow B$ (later, in Chapter 5, we'll see how to measure this probabilistically, so that your credence in $A \supset B$ is always greater than or equal to your credence in the corresponding indicative $A \rightarrow_m B$). But it doesn't follow from this that there are possible worlds in which $A \supset B$ is true and $A \rightarrow_m B$ is false. Part of the lesson of shifting to a non-factualist theory is to avoid exactly this result: we need a finer-grained point of evaluation to model the possibilities relative to which conditionals are true (points that model maximal inferential dispositions).

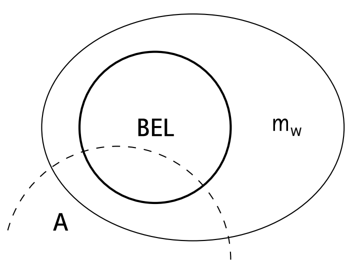
Thus, interpreting "possibility" as "possible world," we block the inference from Strength to the conclusion that there are possibilities in which $A \supset B$ is true and $A \rightarrow_m B$ is false. And if we instead interpret "possibility" in the finer-grained sense, then we avoid the problem by invalidating Weak Sufficiency* since, measuring information gain in terms of finer-grained possibilities, it will turn out that accepting $A \rightarrow_m B$ will involve gaining more information than accepting $A \supset B$. Another way to put the point is this. Indicatives are stronger than material conditionals only in their inferential, non-factual, dimension—that is, only in the constraints on inferential dispositions they impose. But conforming to these inferential dispositions is easy—it only requires coming to believe the material conditional.

Summary: *Indicative bounding resolved*

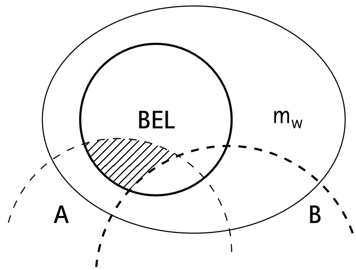
To believe a conditional is to be inferentially disposed in accordance with it; to leave open the possibility of a conditional is to not be inferentially disposed incompatibly with it. The constraints on an agent's factual beliefs and inferential dispositions imposed by an indicative conditional are stronger than those imposed by the corresponding material conditional (which only constrain the agent's factual beliefs). This allows an agent to leave open the possibility of the indicative while believing the material conditional, even though the conditions under which she believes the indicative conditional are satisfied by her merely believing the material conditional, as long as she leaves open the possibility of its antecedent.

I conclude this section with an informal proof that our theory predicts Weak Sufficiency*, Strength, and Conditionalization$^+$:

Predicting Weak Sufficiency*
Suppose **AB** is compatible with your beliefs and m is bounded by your beliefs and that you do not believe $A \rightarrow_m B$. By Conditional Belief, there must be a world w compatible with your factual beliefs such that you are not disposed to infer **B** from $m_w^{\mathbf{A}}$. Because m is bounded by your beliefs and **A** compatible with your beliefs, $\mathbf{BEL} \cap m_w^{\mathbf{A}} \neq \varnothing$:



Thus, by Factual Determination, you are disposed to infer **B** from $m_w^{\mathbf{A}}$ iff your belief state entails $m_w^{\mathbf{A}} \supset \mathbf{B}$. Thus, since you are not disposed to infer **B** from $m_w^{\mathbf{A}}$, your belief state must not entail $m_w^{\mathbf{A}} \supset \mathbf{B}$, which means there are some $\mathbf{A\bar{B}}$-worlds in your belief state:



Furthermore, it follows that ruling out all and only those worlds is sufficient for believing $\mathtt{A} \rightarrow_{\mathrm{m}} \mathrm{B}$. The reason is that once you eliminate all of the $\mathbf{A\bar{B}}$-worlds from your belief state, your new belief state will entail $m_w^{\mathbf{A}} \supset \mathbf{B}$. And since this holds for each w in your belief state, this is sufficient for being disposed to infer **B** from $m_w^{\mathbf{A}}$, for each w in your belief state. And this is sufficient for believing $\mathtt{A} \rightarrow_{\mathrm{m}} \mathrm{B}$.

Predicting Strength

It is possible for an agent to accept $A \rightarrow \neg B$ and not rule out the possibility that $A \supset B$.[8] Suppose our belief state contains $\mathbf{A\bar{B}}$-worlds and $\mathbf{\bar{A}}$-worlds. Then, by the reasoning above, we will believe $\mathtt{A} \rightarrow_{m} \bar{\mathtt{B}}$ (as long as m is bounded by our beliefs). And the fact that there are some $\mathbf{\bar{A}}$-worlds compatible with our beliefs is sufficient to establish that $\mathbf{A} \supset \mathbf{B}$ is compatible with our beliefs.

Predicting Conditionalization^{+}

If Conditionalization^{+} failed, then it would be possible to believe A and also leave open the possibility that $\neg A$. But given that belief and leaving open are duals, this is impossible.

[8] In Chapter 5, we will be able to demonstrate this in a more convincing manner with the probabilities of these conditionals.

3.4 Comparisons with Nearby Theories

The motivation for my contextualist domain inferentialist theory of conditionals comes from our goal of having a unified theory of conditionals that can provide a motivated response to the bounding puzzles. In the next chapter, I offer a formal semantics for contextualist domain inferentialism. However, to finish our motivation for the theory, in this final section, I want to compare the view with two alternatives:

- A factualist theory in which conditionals are indeterminate (Stalnaker 1980, Schulz 2017).
- A non-factualist invariantist theory in which conditionals are not context-dependent (Bradley 2012).

3.4.1 Factualist Indeterminacy

According to this theory, I went wrong in claiming that conditionals encode contents for which there is no fact of the matter whether they are true or false. Instead, according to this view, we should rather think that metasemantics for our language often does not determine a unique interpretation for conditionals—that is, there are often multiple, equally admissible, candidates for the contents of conditionals, and thus multiple, equally admissible (and sometimes conflicting) truth values. This is the view taken by Stalnaker 1980 in his defense of Conditional Excluded Middle.

On Stalnaker's version of this theory, conditionals are evaluated relative to a selection function f that encodes a total closeness relation on the set of all worlds. We will adapt Stalnaker's insight here by defining the conditional's truth in terms of the truth of its consequent at the closest world to w in the conditional's domain at w:

Stalnaker Semantics
$[\![A \to B]\!]^{w,m,f} = 1$ iff $f(w, m_w^{\mathbf{A}}) \in \mathbf{B}$.

The indeterminacy view then holds that English and other natural languages do not determine a unique closeness relation, and thus do not determine a unique proposition expressed by $A \to_m B$. Instead, the propositional content of $A \to_m B$ is indeterminate between the propositions in the following set:

Content Candidates
$\{\mathbf{P} : \exists f : \mathbf{P} = \{w\colon [\![A \to B]\!]^{w,m,f} = 1\}\}$

Suppose Jones has a fair coin but does not flip it. Then the conditional (14) is predicted to be indeterminate because no fact settles which way it would have landed:

(14) If Jones had flipped the coin, it would have landed tails.

Some of its admissible contents will be true (since the f_1-closest flip-world is a tails world) and some of its admissible contents will be false (since the f_2-closest flip-world is a heads world). But since each of these contents is admissible, the conditional is neither determinately true nor determinately false. Since assertively uttering an indeterminate sentence is to assert each of its admissible contents, one mark of lack of determinate truth value is unassertability (cf. Dorr 2003, Sud 2018). Thus, this view correctly predicts that (14) is unassertable in the situation described.

I will discuss two problems for this kind of view.

Problem 1: Indeterminate Probabilities

One problem facing this strategy is that when A is semantically indeterminate, so are sentences like *probably A*. Consider an expression that admits of borderline cases that is plausibly an instance of semantic indeterminacy: perhaps *a lot*. Suppose that we know Smith has some number of books on the borderline of *a lot*: perhaps 100. In such a case, it will be semantically indeterminate whether Smith has a lot of books, and thus the sentence (15) will be unassertable:

(15) #Smith has a lot of books.

However, notice that the following sentences seem equally unassertable (cf. Dorr 2003, Williams 2014):[9]

(16) a. #Smith probably has a lot of books.
b. #It's fifty-fifty whether Smith has a lot of books.
c. #It's unlikely that Smith has a lot of books.

Indeed, a very plausible explanation for the unassertability of (16) is that they are just as indeterminate as (15).[10] Indeed, it seems reasonable to assert that it's indeterminate whether Smith probably has a lot of books. The problem for this strategy is that conditionals behave very differently. A conditional like (14) is predicted to be indeterminate on this theory, and yet embeds felicitously under probability operators:

(17) It is fifty-fifty whether if Jones had flipped the coin, it would have landed tails.

Indeed, it will be crucial to predicting the probabilities of conditionals that differences in their behavior across various sequences are incorporated into the calculation of their probabilities (see Chapter 4). This difference, then, casts doubt on the semantic indeterminacy strategy.

When I give this argument to people, I often hear the following response: (i) the future is open, thus (ii) statements about the future are indeterminate, even though (iii) probability claims about the future are often not indeterminate. I think (iii) is true—notice that the following is true when said of a fair coin that's about to be flipped:

[9] It does seem possible to reinterpret the sentences so that they are assertable. For instance, we might treat them as metalinguistically deferred, as in: "It's unlikely that Smith has a lot of books by academic standards," where the uncertainty is about just what the academic standards for having a lot of books are. But in this alternative interpretation, the embedded sentence is no longer semantically indeterminate. Notice that we could do the same for (15).

[10] Similar results hold for incompletely defined predicates. Suppose I stipulate that natural numbers less than 100 are *noice* and natural numbers greater than 105 are *not noice*. Then, it's semantically indeterminate whether 101 is noice. And, just as (i-a) is unassertable, so is (i-b):

(i) a. #101 is noice.
b. #101 is probably noice.

(18) It's fifty-fifty that it will land heads.

However, I think both (i) and (ii) are questionable. Someone could infer from (iii) that the future is *not* open—that there is a fact about what will happen that we simply do not know (this is the Ockhamist position).[11] Or, someone could infer from (iii) that, while the future is open, it doesn't follow that statements about the future are indeterminate—rather, such claims are non-factual in a way, similar to conditionals, that permits them to have non-trivial probabilities. Or, finally, someone could infer from (iii) that, while statements of the future are indeterminate, whatever sense this is, they must not be *semantically indeterminate*.[12] My argument that semantic indeterminacy projects out of modal embeddings is compatible with any of these responses, and I will not take a stand on which is correct here.

Problem 2: Challenges to Weak Sufficiency

The second problem facing this kind of view is that it struggles to explain what was intuitively appealing about Weak Sufficiency:

> **Weak Sufficiency**
>
> Coming to accept $A \to B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.

This is due in part to the fact that we lack a theory of assertion that extends to indeterminate sentences. Anticipating this concern, Stalnaker 2014 proposes an updated theory of assertion for conditionals (and modal sentences like *might A*) in which assertively uttering a sentence is a proposal to minimally change the context so that the proposition expressed by that sentence is accepted in the new context set:[13]

[11] See Pruss 2010 for an argument of this kind.

[12] Perhaps they are metaphysically indeterminate, as in Williams 2008c; Barnes & Cameron 2009, 2011; Barnes & Williams 2011.

[13] The context set of a context is the set of worlds compatible with what's mutually presupposed by the participants in that context.

Prospective Assertion
An utterance of $A \rightarrow B$ is a proposal to minimally change the context set to c' such that c' entails $[\![A \rightarrow B]\!]^{m_{c'}, f}$ for all fs.

In other words, what this says is that, for conditionals (and other modal sentences), assertion is *prospective*—it proposes a minimal change to what's presupposed in the context such that, relative to the new context set, the conditional is determinately true throughout. This predicts that the minimal change to the common ground proposed by an assertion of $A \rightarrow_m B$ is to eliminate every $\mathbf{A\bar{B}}$-world (that is, to come to believe $\mathbf{A} \supset \mathbf{B}$). Doing that is enough to ensure that the only **A**-worlds in c' are **B**-worlds, and hence that all of the closest **A**-worlds to any in c' will be **B**-worlds, and that is enough to ensure that the conditional will be true at every world in c'.

So, opting for Prospective Assertion for conditionals allows Stalnaker to predict something like Weak Sufficiency while also predicting Strength—since, relative to any selection function f, $A \rightarrow B$ may be false even though $A \supset B$ is true (this will hold at some $\mathbf{\bar{A}}$-worlds, but possibly at different ones for different selection functions).

I will limit myself to two remarks about this strategy.[14] The first is that merely stipulating Prospective Assertion is unsatisfying and ad hoc. Ideally, we would like some independent motivation for that claim, other than that it gets the right results for conditionals (and perhaps also modals; see Mandelkern 2020 for discussion of the latter). Notice that Stalnaker's early theory of assertion fares better in this regard. On the earlier theory, assertion is a proposal to coordinate our beliefs on a piece of information. This information is what is believed by the speaker, and what the speaker aims to be common ground between her and her audience. This motivates the view of assertion whereby it is a proposal to intersect the information state of the context (modeling the information compatible with the common ground of the context) with the asserted proposition. On the Prospective Assertion theory, assertion is instead a proposal to change

[14] A similar strategy is employed by Schulz 2017 to handle his version of the Subjunctive Bounding Puzzle. My remarks here apply to Schulz as well—see Khoo 2020 for further discussion.

the context minimally so that the uttered sentence comes out true in the new context. It is less clear how to motivate this view in terms of information exchange.

The second concern is that the proposal predicts that you cannot come to learn $[\![A \to B]\!]^c$ by accepting an utterance of $A \to B$ in c, which is surprising. This is because what you are instead predicted to learn is $[\![A \to B]\!]^{c'}$—the prospective content of the conditional. I take this to be a mark against the theory, since intuitively what we wanted to learn via an assertive utterance of $A \to B$ was its content, but instead we learn some distinct content, which is what the sentence would express were the utterance accepted.

By contrast, my non-factualist theory avoids both concerns. My theory of assertion is independently motivated: an assertion of a conditional is a proposal to accept its content—to be inferential disposed in accordance with it—and, as we saw, to be so inferentially disposed just is to minimally update your factual beliefs so that they determine the relevant inferential dispositions. Hence, we predict that you can come to learn the content of $A \to B$ relative to the context in which it is uttered.

3.4.2 Non-Factualist Invariantism

On the other side of things, some readers may wonder about the contextualist part of my theory. Why not opt for a more familiar non-factualist invariantism? After all, typically, non-factualist theories have been proposed exactly to handle subjective language (in this case, epistemic modal vocabulary) without running afoul of standard objections to contextualism stemming from disagreement and speech reports.

We have already seen a version of this view (due to Mellor 1993, but see also Bradley 2012), which I labeled Simple Inferentialism:

> **Simple Inferentialism**
> The contents of conditionals encode the disposition to infer their consequents from their antecedents.

Such a view promises to predict Or to If and Weak Sufficiency without running into the indicative bounding puzzle, and furthermore it requires no additional detour through bounded modal bases, since it doesn't relativize conditionals to modal bases at all (see also Santorio & Goldstein 2021, Santorio forthcoming).

Above, I already sketched my initial motivation against such a view: since Simple Inferentialism is implausible for subjunctive conditionals, we need another account of their meaning. Thus, adopting Simple Inferentialism for indicative conditionals forces us to accepted a disunified view of indicative and subjunctive conditionals. Furthermore, it leaves us with no explanation of the non-factuality of certain subjunctive conditionals (indeed, on Gibbard 1981's bifurcated theory, it is only indicatives that are non-factual). Given the parallels between their logics and the similarities in the puzzles that arise for them, this is an unhappy result.[15]

Relativizing conditionals to modal bases allows for a unified theory (via Domain Inferentialism), but why think the modal base is initialized by the context? One reason is that, while subjunctive conditionals are characteristically metaphysical and indicatives characteristically epistemic, in some contexts, there are subjunctives that are epistemic and indicatives that are metaphysical. I argue for this in Chapter 8. This suggests that context is playing a non-trivial role in assigning modal bases to these conditionals.

Another reason comes from Sly Pete cases. The classic case is due to Gibbard 1981:[16]

[15] In addition to the many features they share in common regarding their logics (see Bennett 2003: 172–176), Swanson 2011b observes biscuit and "stand off" (Sly Pete cases—see below) conditionals in both indicatives and subjunctives. Biscuit conditionals are so-called because of the famous example in Austin 1970: 158:

(i) There are biscuits on the sideboard if you want them.

I do not give any treatment of biscuit (or relevance) conditionals. See Osborn 1965, DeRose & Grandy 1999, Siegel 2006, Franke 2007, Ebert et al. 2008, Predelli 2009, Krifka 2014, Francez 2015 for further discussion.

[16] See Williams 2008b, Stalnaker 2011 for similar arguments involving the Sly Pete case. Stalnaker adds a variation on the case (Stalnaker 2011: 245–247) that motivates distinguishing the domain of an epistemic indicative conditional from the information available in the common ground of the context. I lack the space to fully engage with Stalnaker's variant here.

Sly Pete. Sly Pete and Mr. Stone are playing poker. Pete is an unscrupulous cheater: he has two of his henchmen, Zack and Jack, planted to help him win. Stone has placed bets on the current hand, and it is now up to Pete to call or fold. Zack has seen Stone's hand, which is quite good, and signaled its contents to Pete. Jack has seen both hands, and seen that Pete has the losing hand. Stone, suddenly suspicious that Pete is cheating, demands that the room be cleared. Zack heads to a nearby pool hall, while Jack goes home. It is now after the hand has been played, but before the outcome is known. Zack says:

(19) If Pete called, he won.

At the same time, Jack says:

(20) If Pete called, he lost.

Here are two seemingly conflicting intuitions about this case:

- **No Conflict**: Both Zack and Jack's claims are true.
- **Conjoined Conflict**: The sentence (21) is necessarily false:

(21) #If Pete called, he won, and if Pete called, he lost.

On the face of it, it might seem that these intuitions cannot both be right. After all, if Zack and Jack's claims are both true, and what Zack said just is that if Pete called, he won and what Jack said is that if Pete called, he lost, then it should follow that any sentence that conjoins what was said by each (like (21)) should be consistent. I will come back to this point in a moment, when I argue that endorsing context sensitivity allows us to avoid this apparent clash.

For now, let us bolster the case for both intuitions. Suppose, contra **No Conflict**, that one of Zack or Jack's claims is false. If that were so, there should be a fact that falsifies that claim (whichever it is). Now, one might think that the fact that Pete has the weaker hand falsifies Zack's claim. However, DeRose 2010 argues that this thought is mistaken, for the following reason. Notice that Zack has seen Stone's hand, and sees that it is quite good. Perhaps he even has a very high credence that Pete has the weaker hand. If this fact falsified Zack's claim that if Pete called, he won, then it should be incoherent for Pete to believe what he said. After all, in general, if I believe A is likely and A entails $\neg B$, then I can't also rationally believe B. Yet, it seems that Zack can believe it is likely that Pete has the weaker hand and still rationally believe what he said. Roughly, he reasons as follows: if Pete called, he must have had the higher hand, since, after all, he knew both hands and wanted to win. But, then, if the fact that Pete had the weaker hand does not falsify Zack's belief, it's hard to see what else could. Furthermore, it is equally hard to see what fact could falsify Jack's belief. So, **No Conflict** seems to be on secure ground.

What about **Conjoined Conflict**? Notice first that (21) just seems incoherent: it is hard to think of what someone uttering it might be trying to communicate. Furthermore, like with any contradiction, even uttering the second conjunct under an epistemic possibility modal is infelicitous:

(22) #If Pete called, he won; and maybe if Pete called, he lost.

Thus, I think there are good reasons to take **Conjoined Conflict** seriously as well.

Opting for context sensitivity allows us to avoid the apparent clash between these two intuitions. On a contextualist theory, the content of Zack's conditional is generated by Zack's information, which entails the material conditional *Pete called* $\supset$ *Pete won*, whereas the content of Jack's conditional is generated by Jack's information, which entails *Pete called* $\supset$ *Pete lost*. Then, at the actual world, both Zack and Jack's conditionals will be true.[17] However, the conjunction of their

[17] Indeed, one can even accept both. Here is a simple case:

- $\mathbf{BEL}_Z = \{w_1, w_2\}, \mathbf{BEL}_J = \{w_2, w_3\}$
- $z_{w_1} = z_{w_2} = z_{w_3} = \{w_1, w_2\}; j_{w_1} = j_{w_2} = j_{w_3} = \{w_2, w_3\}$
- $\mathbf{P} = \{w_1, w_3\}, \mathbf{W} = \{w_1\}$

Zack accepts $P \rightarrow_z W$ since for all $w \in \mathbf{BEL}_Z : \mathbf{P} \cap z_w \subseteq \mathbf{W}$. But still, Zack can accept Jack's claim $P \rightarrow_j L$, as long as he updates his belief state to remove the sole **P**-world from it. Notice that Zack regards Jack's informational modal base as non-evidential (and hence not bounded by his belief state), but still once w_1 is gone from Zack's belief state, conditionals (21) cannot express a true proposition at any context, hence it is judged incoherent. No one can accept both $P \rightarrow_{m_c} W$ and $P \rightarrow_{m_c} L$ since they place incompatible constraints on the agent's inferential dispositions.

By contrast, an invariantist view cannot predict this. According to it, what Zack and Jack say are incompatible—what Zack believes just is the contradictory of what Jack does, according to such a view. Therefore, it cannot be the case that both of their beliefs are correct (although it may be that both had a right to assert what they did). So, such a view does not predict No Conflict.

Indeed, even fancy dynamic theories cannot predict this result, as long as the truth predicate is deflationary. According to such theories, to accept a conditional $A \rightarrow B$ is for your information state to be compatible with A and entail $A \supset B$. As such, no information state accepts both $A \rightarrow B$ and $A \rightarrow \neg B$—this correctly predicts Conjoined Conflict. But for the same reason no information state accepts both that Zack's claim is true and that Jack's claim is true (given that the truth predicate is deflationary).[18]

I argued for three claims in this chapter, repeated here, with the arguments for them summarized briefly:

<u>Claim 1</u>: Conditionals—both indicative and subjunctive—encode constraints on inferential dispositions.

- Argument: this is independently plausible for indicative conditionals, and extending it to subjunctives allows for a unified theory that can resolve the bounding puzzles.

<u>Claim 2</u>: What it is to believe or leave open the possibility of a conditional is a matter of an agent's inferential dispositions, where these are entirely determined by her factual beliefs.

he will be disposed in accordance with $P \rightarrow_j L$. This is because w_2 is such that $j_{w_2} \cap \mathbf{P} \subseteq \mathbf{L}$ and so Zack will have the disposition trivially.

[18] We could give up deflationism and hold that:

(i) Zack's claim is true iff Zack's information entails X (where this is the sentence he uttered by that claim).

But this view then predicts, incorrectly, that it should be coherent to say:

(ii) #Zack's claim is true, but I don't accept it.

- Argument: this captures the acceptance conditions of conditionals in a way that distinguishes them from material conditionals and that explains the contrast between cases of resolvable and unresolvable ignorance.

<u>Claim 3</u>: The above claims provide an independent motivation for rejecting Conditonalization$^-$ and Weak Sufficiency, while at the same time explaining what was compelling about them in the first place, thus avoiding the bounding puzzles.

- Argument: the resulting theory motivates the rejection of Conditionalization$^-$ since it takes more to learn a subjunctive conditional $A \Rightarrow B$ than simply ruling out all possibilities in which it is factually false; it also motivates the rejection of Weak Sufficiency understood as the claim that there is nothing more to believing $A \rightarrow B$ than believing $A \supset B$—in fact, there is more to this, and it involves coming to have certain inferential dispositions. It is just that having the relevant inferential dispositions doesn't require learning anything factually stronger than the corresponding material conditional $A \supset B$.

I turn now to providing a truth-conditional semantics that implements Domain Inferentialism formally.

4

Sequence Semantics

So far, we have described the meanings of conditionals mostly informally, in terms of the inferential dispositions they encode. In this chapter, I aim to do two things. One is to show how to connect Domain Inferentialism with the formal semantic framework pioneered by Stalnaker (1968) and refined by van Fraassen (1976), McGee (1989), and Bradley (2012). The second is to show how we can embed the insights of the previous chapter in a formal semantics for conditionals that can generate precise predictions about the logic of conditionals, as well as their probabilities and acceptance conditions.

Our jumping-off point will be Stalnaker 1968. Roughly, according to Stalnaker's early theory, an indicative conditional $A \rightarrow B$ is true iff the closest epistemically accessible A-world is a B-world; and a subjunctive conditional $A \Rightarrow B$ is true iff the closest metaphysically accessible A-world is a B-world. This theory captured three semantic features of conditionals that I take to be essential: that indicative and subjunctive conditionals receive a uniform semantics, that indicatives are stronger than material conditionals, and that subjunctives are weaker than strong subjunctives.[1]

I think this kind of semantics is exactly right. But I think we haven't yet drawn the right lesson from it. After all, as we saw in Chapter 3, such a semantics predicts that both Weak Sufficiency and Strong Sufficiency are false:

[1] Indicatives can be stronger than material conditionals. Suppose A is false; then it is not automatically the case that the closest epistemically accessible **A**-world is a **B**-world, in which case the indicative $A \rightarrow B$ may be false even though its material counterpart $A \supset B$ is true. Subjunctives can be weaker than strong subjunctives. Suppose that among the metaphysically accessible **A**-worlds are both **B** and $\bar{\mathbf{B}}$-worlds. Then, the strong subjunctive $A \mathrel{\Box\!\!\rightarrow} B$ will be false. But $A \Rightarrow B$ may be true, as long as the closest metaphysically accessible **A**-world is a **B**-world.

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0005

Weak Sufficiency

Coming to accept $A \to B$ (when you regard $A \wedge B$ as possible) is to come to believe nothing more than that $A \supset B$ is true.

Strong Sufficiency

If you are sure that A is false, then you accept $A \Rightarrow B$ only if you believe $A \,\Box\!\!\to B$.

Take Weak Sufficiency and suppose that among your belief worlds are some $\bar{\mathbf{A}}$-worlds where $A \to B$ is false. Then, coming to accept $A \to B$ is to rule out these worlds, and thus learn something more than $A \supset B$. Take Strong Sufficiency and suppose that among your belief worlds are some where $A \,\Box\!\!\to B$ is false but $A \Rightarrow B$ is true. Then, you can accept $A \Rightarrow B$ by ruling out worlds where it is false, and still not accept $A \,\Box\!\!\to B$.

Nonetheless, I think we can appreciate the insights of Stalnaker's theory by combining his compositional semantics with our theory that conditionals encode inferential dispositions. In this first part of the chapter, I show how to do that.

4.1 From Selection Functions to Sequences

Stalnaker makes use of a selection function f that maps a world w and a proposition $\mathbf{A}$ to the closest $\mathbf{A}$-world to w. I want to generate the same result in a slightly different way, by collapsing the role of the world and selection function, and letting both be played by a sequence of worlds, σ (cf. van Fraassen 1976, Bradley 2012, Bacon 2015, Santorio forthcoming). This will enable us to more simply formulate the kind of fine-grained distinctions between possibilities generated by conditional contents, although strictly speaking we could also make the same points in a selection functional framework.

A **sequence** of worlds σ is an ordered n-tuple of worlds in which each world in $\mathbf{W}$ appears exactly once. Let Ω be the set of all sequences of worlds, and define:

Sequences

i. σ_w is the first world in σ.
ii. $\sigma(\mathbf{A})$ is the first $\mathbf{A}$-world in σ.
iii. $w >_\sigma w'$ iff w comes before w' in σ.
iv. $\sigma(\mathbf{A})$ is undefined if $\mathbf{A} = \varnothing$.

We can define a Stalnakerian selection function in terms of a sequence as follows:

Selection Function

$$f_\sigma(w, \mathbf{A}) = \begin{cases} w & \text{if } w \in \mathbf{A} \\ \sigma(\mathbf{A}) & \text{otherwise} \end{cases}$$

The domain of a conditional is determined by a modal base m, which is a function from a world w to a set of worlds m_w.[2] We will sometimes relativize modal bases to sequences, and when we do, it will be the first world of the sequence that serves as the input to the modal base function: hence, $m_\sigma = m_{\sigma_w}$.

Given this, we now have two parameters that the truth of a sentence is relativized to: a modal base m and a sequence σ:

$$[\![A]\!]^{\sigma,m} = 1 \text{ iff } A \text{ is true relative to } m, \sigma.$$

I'll come back later to say more about this relation and what the contents of conditionals are on this model. But, for now, we can state our baseline unified semantics of conditionals as follows:

Baseline Semantics

$$[\![A \rightarrow B]\!]^{\sigma,m} = 1 \text{ iff } \sigma(m_\sigma^{\mathbf{A}}) \in \mathbf{B}$$

This states that $A \rightarrow B$ is true at m, σ iff the first $m_\sigma^{\mathbf{A}}$-world in the sequence σ is a $\mathbf{B}$-world. One is welcome to think about σ as providing a kind of closeness relation for a world (its starting world), and, thus understood, our baseline semantics agrees with Stalnaker that

[2] In Chapter 8, I argue that we need two parameters to determine the domain of a conditional—a modal base and an ordering source, as in Kratzer 1981.

an indicative/subjunctive $A \to B$ is true iff the closest **A**-world in its domain is a **B**-world. Any semantic or pragmatic differences between indicatives and subjunctives on this semantics are entirely due to differences in their domains.

I assume that we interpret conditionals (and all modals in fact) relative to factive modal bases:

Factivity
m is factive iff $\forall w : w \in m_w$.

Given this, we immediately predict the following:

Modus Ponens
$A, A \to B \vDash B$

Suppose A and $A \to B$ are true at σ, m. Then, $\sigma(m^{\mathbf{A}}_{\sigma}) \in \mathbf{B}$. And by Factivity, σ_w is an **A**-world. Therefore, σ_w is a **B**-world.[3]

Strong Centering
$A \vDash (A \to B) \equiv B$

Suppose A is true at σ, m. Then by Modus Ponens, we know that if $A \to B$ is true at σ, m, so is B. In reverse: suppose B is also true at σ, m. Then, σ_w is an **AB**-world. So, $\sigma(m^{\mathbf{A}}_{\sigma}) \in \mathbf{B}$. Thus, $A \to B$ is true at σ, m.

Conditional Excluded Middle
$\vDash (A \to B) \vee (A \to \neg B)$

Suppose that $A \to B$ is false at σ, m. It follows that $\sigma(m^{\mathbf{A}}_{\sigma}) \notin \mathbf{B}$. But that is sufficient for $A \to \neg B$ to be true at σ, m. The reasoning from the falsity of $A \to \neg B$ to the truth of $A \to B$ is analogous.

All of these logical principles seem to hold for both indicative and subjunctive conditionals. Modus Ponens seems valid for simple con-

[3] We will see fit to revise this prediction later, in Chapter 5, when we decompose the meaning of $\to$ into an index-shifting component and a modal component.

ditionals (those containing no modals or conditionals in their consequents).[4] Its validity correctly predicts the observation that it is contradictory to say or believe:

(1) a. #If John is not in Athens, he is in Barcelona. And he's in neither: he is in Chicago.
b. #If John hadn't been in Athens, he would have been in Barcelona. And he's in neither: he's in Chicago.

What about Strong Centering? Although perhaps more controversial than Modus Ponens, it, too, strikes me as very plausibly valid. For one, it correctly predicts that it is contradictory to say or believe:

(2) a. #John is in Barcelona, but it is not the case that if he is not in Athens, he is in Barcelona.
b. #John is in Barcelona, but it is not the case that if he hadn't been in Athens, he would have been in Barcelona.

A second mark in favor of Strong Centering comes from betting intuitions about conditionals. Suppose you bet me that if John rolled a prime, he rolled an even. If in fact John rolled on a 3 or a 5, then I win the bet, but if he rolled a 2, you win the bet (set aside for now what happens if he rolled a non-prime). These intuitions are evidence that the conditional (3) is true if John rolled an even prime, and false if he rolled an odd prime.

(3) If John rolled a prime, he rolled an even.

McDermott 2007 raises a potential counterexample to Strong Centering for subjunctive conditionals. Suppose a fair coin is tossed twice and I bet it will land heads both times, and it does. Nonetheless, it seems incorrect to say:

[4] The purported counterexamples to Modus Ponens all involve such complex embeddings: see McGee 1985, Kratzer 1986, Gillies 2004, Khoo 2013, Khoo & Mandelkern 2019. We will return to the extra complexity introduced by such conditionals in Chapter 5.

(4) If the coin had landed heads at least once, I would have won.

However, this case is complicated for a number of reasons, illustrating why it is hard to directly assess the validity of Strong Centering. When you know that you won the bet, saying (4) violates the Maxim of Quantity, since you could have said something relevantly stronger; hence, we expect (4) to be infelicitous in such cases. But when you don't know that you've won the bet, you allow for possibilities in which the coin did land heads exactly once, in which you didn't win the bet (and thus, by Modus Ponens, (4) is false). And so we expect an utterance of (4) to be infelicitous in that case as well. Finally, note that this explanation of the infelicity of (4) is compatible with the conditional being true.

Finally, what about Conditional Excluded Middle? This is perhaps the most controversial of the three principles, ruling out the possibility that $A \to B$ and $A \to \neg B$ could both be false together. The standard purported counterexamples involve pairs of such conditionals, neither of which seem true:

(5) a. If Smith flipped the coin, it landed heads.
 b. If Smith flipped the coin, it landed tails.

Given that the coin was fair, we might be inclined to think that both (5-a) and (5-b) are false. However, recently, many theorists have pushed back on these intuitions in favor of Conditional Excluded Middle.[5] For instance, Moss 2013 observes that, just as (5-a) and (5-b) are unassertable, so are their wide scope negations, which is surprising if (5-a)/(5-b) are both false:

(6) a. It's not the case that if Smith flipped the coin, it landed heads.
 b. It's not the case that if Smith flipped the coin, it landed tails.

By contrast, (7) is assertable, and seems true:

[5] See in particular Stalnaker 1980, DeRose 1999, Cross 2009, Williams 2010, Swanson 2011a. Although most of the recent discussion has been over counterfactuals, the same lessons apply to indicative conditionals.

(7) It is 50-50 likely that if Smith flipped the coin, it landed heads.

Yet, if (5-a) were false merely because it is possible that Smith flipped the coin and it landed tails, we expect the judgments here to be reversed. A more promising explanation of what is going on, then, is that (5-a)–(6-b) are all unassertable because we do not know how the coin landed if Smith indeed flipped it, but (7) is true because it is an accurate description of the probability of (5-a). However, this explanation relies on Conditional Excluded Middle: what we are ignorant about is which of (5-a)/(5-b) is true (I assume here that *heads* = *not tails*).

Further evidence in favor of Conditional Excluded Middle comes from the observation that denials of conditionals seem like affirmations of their CEM counterparts:

(8) A: If Smith is not in Athens, he is in Barcelona.
B: That is false; if he is not in Athens, he is not in Barcelona either.

Relatedly, Conditional Excluded Middle allows us to predict the behavior of quantified conditionals like (cf. Higginbotham 1996, 2003; Klinedinst 2010):

(9) a. No student will succeed if they goof off.
b. Every student will fail if they goof off.

These conditionals seem equivalent, and indeed they are if Conditional Excluded Middle is valid. And, finally, Conditional Excluded Middle gains some support from the fact that it, together with Modus Ponens, entails Strong Centering, which we have already seen is independently plausible.[6] Thus, it seems that the intuitive evidence does favor Conditional Excluded Middle quite strongly.[7]

[6] Suppose A is true and B is false. Then by Modus Ponens it must be that $A \rightarrow B$ is false. Suppose A is true and B is true. Then, by Modus Ponens, $A \rightarrow \neg B$ is false. And then by Conditional Excluded Middle, $A \rightarrow B$ is true.

[7] Note that, even Lewis, who argued that Conditional Excluded Middle fails for counterfactuals, admitted that this principle was intuitively plausible (Lewis 1973a: 79–80).

4.2 Refined Contents and Inferential Dispositions

We have motivated a semantics for conditionals in which their extensions are relativized to a sequence of possible worlds. Now we are in a position to unify this view with Inferential Dispositionalism. Begin with the assumption that the content of a conditional (relative to modal base m) is the set of sequences at which it is true:

$$[\![A \rightarrow B]\!]^m = \{\sigma : \sigma(m^{\mathbf{A}}_{\sigma}) \in \mathbf{B}\}$$

We now need to say what kind of content a set of sequences is, and what it is to believe such a content. I propose that a sequence of possible worlds represents a way things could have been and a learning plan, which is a maximally specific pattern of inferential dispositions. The sequence $\langle w_1, w_2, w_3, w_4 \rangle$ represents things as being as they are at w_1 and the following learning plan:[8]

- Infer $\{w_1\}$ from $\{w_1, w_2, w_3, w_4\}$
- Infer $\{w_2\}$ from $\{w_2, w_3, w_4\}$
- Infer $\{w_3\}$ from $\{w_3, w_4\}$

Thus, two sequences that differ only in their tails, such as $\langle w_1, w_2, w_3, w_4 \rangle$ and $\langle w_1, w_3, w_2, w_4 \rangle$, do not differ factually (they both determine the same factual way things could have been, viz. w_1). Instead, they differ only in terms of the non-factual component of their learning plans. And inferential dispositions are not ways things could have been—while there is a fact of the matter about whether grass is green, there is no fact of the matter about being disposed to infer **B** upon believing **A** (though there may be facts about whether one ought to infer **B** from **A**). Thus, such distinctions between sequences are entirely non-factual.

Now, just as a possible world is a way things could have been, and a set of possible worlds (a possible worlds proposition) is a property

[8] The plan also includes trivial inferential dispositions where what you're inferring is entailed by what it's being inferred from: infer $\{w_1\}$ from $\{w_1\}$ and infer $\{w_1, w_2\}$ from $\{w_1\}$, and so on.

of how things could have been, a set of sequences is a property of how things could have been and learning plans. I call such sets **refined contents**, and I'll use uppercase teletype letters $\mathtt{A}$, $\mathtt{B}$, $\mathtt{C}, \ldots$ to denote them. A refined content models a non-maximal set of inferential dispositions. For instance, the set $\{\langle w_1, w_2, w_3, w_4\rangle, \langle w_1, w_2, w_4, w_3\rangle\}$ encodes the inferential dispositions:

- Infer $\{w_1\}$ from $\{w_1, w_2, w_3, w_4\}$
- Infer $\{w_2\}$ from $\{w_2, w_3, w_4\}$
- Infer $\{w_3, w_4\}$ from $\{w_3, w_4\}$

Each refined content also encodes a propositional content, given by the set of worlds that appear first in any of its members. So, the propositional content of $\{\langle w_1, w_2, w_3, w_4\rangle, \langle w_1, w_2, w_4, w_3\rangle\}$ is $\{w_1\}$. The coarsening operator $\Downarrow$ maps a refined content to the propositional content it encodes:

Coarsening
$\Downarrow\mathtt{A} = \{w : \exists\sigma \in \mathtt{A} : \sigma_w = w\}$

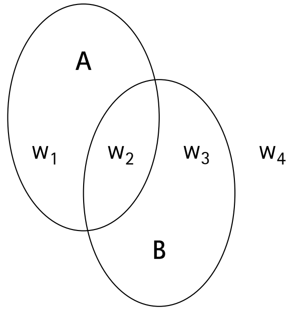
I'll use '$\mathbf{A} \mapsto \mathbf{B}$' as shorthand for the disposition to infer $\mathbf{B}$ from $\mathbf{A}$. Just as propositions are properties of possible worlds, inferential dispositions are properties of sequences:

$$\mathbf{A} \mapsto \mathbf{B} = \{\sigma : \sigma(\mathbf{A}) \in \mathbf{B}\}$$

A refined content encodes an inferential disposition iff it entails it:

Encoding
A refined content $\mathtt{X}$ encodes inferential disposition $\mathbf{A} \mapsto \mathbf{B}$ iff $\mathtt{X} \subseteq \mathbf{A} \mapsto \mathbf{B}$.

Thus, to check to see whether some refined content $\mathtt{X}$ encodes some inferential disposition $\mathbf{A} \mapsto \mathbf{B}$, we need to check whether, for each sequence in $\mathtt{X}$, its first $\mathbf{A}$-world is a $\mathbf{B}$-world. Take a simple model with four worlds w_1, w_2, w_3, w_4 and suppose that $\mathbf{A}$ is true at w_1, w_2 and $\mathbf{B}$ is true at w_2, w_3:



Let's look at two refined contents X and Y:

$$\mathtt{X} = \begin{Bmatrix} \langle w_3, w_2, w_4, w_1 \rangle \\ \langle w_3, w_2, w_1, w_4 \rangle \end{Bmatrix}$$

$$\mathtt{Y} = \begin{Bmatrix} \langle w_3, w_2, w_1, w_4 \rangle \\ \langle w_3, w_1, w_2, w_4 \rangle \end{Bmatrix}$$

To start, notice that each encodes the same factual information, $\{w_3\}$. But X also encodes the inferential disposition $\mathbf{A} \mapsto \mathbf{B}$, since at each of its sequences, the first **A**-world is a **B**-world:

First **A**-world is a **B**-world

$$\mathtt{X} = \begin{Bmatrix} \langle w_3, w_2, w_4, w_1 \rangle \\ \langle w_3, w_2, w_1, w_4 \rangle \end{Bmatrix}$$

First **A**-world is a **B**-world

By contrast, Y does not encode $\mathbf{A} \mapsto \mathbf{B}$, since, although the first **A**-world of its first sequence is w_2, which is a **B**-world, the first **A**-world of its second sequence is w_1, which is a $\bar{\mathbf{B}}$-world:

First **A**-world is a **B**-world

$$\mathtt{Y} = \begin{Bmatrix} \langle w_3, w_2, w_1, w_4 \rangle \\ \langle w_3, w_1, w_2, w_4 \rangle \end{Bmatrix}$$

First **A**-world is a $\bar{\mathbf{B}}$-world

Next, just as refined contents encode both factual information and inferential dispositions, the two components may interact. Thus, consider the distinction between the following refined contents:

$$\mathtt{P} = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3, w_4 \rangle & \langle w_2, w_1, w_3, w_4 \rangle \\ \langle w_1, w_2, w_4, w_3 \rangle & \end{array} \right\}$$

$$\mathtt{Q} = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3, w_4 \rangle & \\ \langle w_1, w_2, w_4, w_3 \rangle & \langle w_2, w_1, w_4, w_3 \rangle \end{array} \right\}$$

These two contents agree on their factual information (w_1, w_2) as well as their inferential dispositions. Both encode:

- $\mathbf{W} \mapsto \mathbf{A}$

And, yet they differ still in how their factual and inferential components interact. In particular, they encode different inferential dispositions at different worlds. At w_1, neither encodes the disposition to infer either $\mathbf{B}$ or $\bar{\mathbf{B}}$ from $\bar{\mathbf{A}}$. But at w_2, $\mathtt{P}$ encodes the disposition to infer $\mathbf{B}$ from $\bar{\mathbf{A}}$, and at w_2, $\mathtt{Q}$ encodes the disposition to infer $\bar{\mathbf{B}}$ from $\bar{\mathbf{A}}$:

First $\bar{\mathbf{A}}$-world is a $\mathbf{B}$-world

$$\mathtt{P} = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3, w_4 \rangle & \langle w_2, w_1, w_3, w_4 \rangle \\ \langle w_1, w_2, w_4, w_3 \rangle & \end{array} \right\}$$

$$\mathtt{Q} = \left\{ \begin{array}{ll} \langle w_1, w_2, w_3, w_4 \rangle & \\ \langle w_1, w_2, w_4, w_3 \rangle & \langle w_2, w_1, w_4, w_3 \rangle \end{array} \right\}$$

First $\bar{\mathbf{A}}$-world is a $\bar{\mathbf{B}}$-world

Thus, we need another way to distinguish them. A targeted inferential disposition encodes both a factual and inferential constraint, demanding the inference from $\mathbf{A}$ to $\mathbf{B}$ if at world w and imposing no constraints on inferential dispositions otherwise. We will appeal to targeted inferential dispositions to characterize refined contents.

$$\mathbf{A} \mapsto_w \mathbf{B} = \{\sigma\text{: if } \sigma_w = w\text{, then } \sigma(\mathbf{A}) \in \mathbf{B}\}.$$

We can understand $\mathbf{A} \mapsto_w \mathbf{B}$ intuitively by glossing it as "infer **B** from **A** at w." Now, we can distinguish $\mathtt{P}$ and $\mathtt{Q}$ by their targeted inferential dispositions:

- $\mathtt{P}$ encodes $\bar{\mathbf{A}} \mapsto_{w_2} \mathbf{B}$
- $\mathtt{Q}$ encodes $\bar{\mathbf{A}} \mapsto_{w_2} \bar{\mathbf{B}}$

Finally, some inferential dispositions are trivial, in the sense that they are encoded by every refined content:

- $\mathbf{A} \mapsto \mathbf{B}$ is trivial if $\mathbf{A} \subseteq \mathbf{B}$.

Others are non-trivial, in the sense that they are encoded only by proper subsets of the set of all sequences Ω:

- $\mathbf{A} \mapsto \mathbf{B}$ is non-trivial if $\mathbf{A} \nsubseteq \mathbf{B}$ and $\mathbf{A} \nsubseteq \bar{\mathbf{B}}$.

We can now verify that our semantic theory predicts the inferential dispositions encoded by a conditional as discussed in Chapter 3:

Conditional Inferential Dispositions
A conditional $A \rightarrow_m B$ encodes the disposition $m_w^{\mathbf{A}} \mapsto_w \mathbf{B}$, for each w.

Discussion. Given our baseline semantics, $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is the set of sequences $\{\sigma : \sigma(m_\sigma^{\mathbf{A}}) \in \mathbf{B}\}$. And remember that the targeted disposition $m_w^{\mathbf{A}} \mapsto_w \mathbf{B}$ just is that set of sequences $\{\sigma : \sigma(m_w^{\mathbf{A}}) \in \mathbf{B}\}$. But now take any w, such that there is a sequence σ whose first world is w, such that $\sigma(m_\sigma^{\mathbf{A}}) \in \mathbf{B}$. Since σ's first world is w, we immediately have that $\sigma(m_w^{\mathbf{A}}) \in \mathbf{B}$, and thus that σ is a member of the targeted disposition $m_w^{\mathbf{A}} \mapsto_w \mathbf{B}$.

Crucially, note that this follows from Baseline Semantics and our interpretation of refined contents.

I want to conclude this section by noting that we can define two distinct truth predicates of refined contents in this framework. We began with the truth of a refined content relative to a sequence—call this "thin truth":

Thin Truth
$\mathtt{A}$ is true relative to σ iff $\sigma \in \mathtt{A}$.

But we can now also define the factual truth of a refined content relative to a world w; factual truth at w is truth at all sequences that share w as a first member:

Factual Truth

$$\mathtt{A} \text{ is } \begin{cases} \text{factually true at } w & \text{if} \quad \forall \sigma : \sigma_w = w : \sigma \in \mathtt{A} \\ \text{factually false at } w & \text{if} \quad \forall \sigma : \sigma_w = w : \sigma \notin \mathtt{A} \\ \text{non-factual at } w & \quad\quad \text{otherwise} \end{cases}$$

A refined content is factually true when its truth is settled entirely by the truth of its corresponding propositional content. Given our truth conditions for conditionals, it follows that:

$$\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B} \text{ is } \begin{cases} \text{factually true at } w & \text{if} \quad w \in \mathbf{AB} \text{ or } m_w^{\mathbf{A}} \subseteq \mathbf{B} \\ \text{factually false at } w & \text{if} \quad w \in \mathbf{A\bar{B}} \text{ or } m_w^{\mathbf{A}} \subseteq \mathbf{\bar{B}} \\ \text{non-factual at } w & \quad\quad \text{otherwise} \end{cases}$$

Factual truth is a useful notion to have in hand, and it shows how our semantics does predict a strong sense in which many conditionals are non-factual. Furthermore, as we'll see, our notion of non-factuality does not prevent us from believing or leaving open the possibility of conditionals. In particular, although we cannot believe a conditional we believe to be non-factual, we can leave open the possibility of such a conditional. And, we can believe an indicative conditional even if we leave open the possibility that it is non-factual. We'll see how this works below. First, we need show how to model the factual determination of inferential dispositions (which we do in §4.3), and how to model refined belief/leaving open (which we do in §4.4).

Summary: ***Sequences and inferential dispositions***

Contents encoding inferential dispositions (refined contents) can be modeled as sets of sequences of possible worlds, thus unifying Domain Inferentialism and sequence semantics.

4.3 Factual Determination

Recall Factual Determination from Chapter 3:

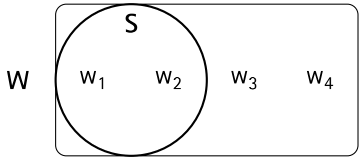
Factual Determination
An agent S is disposed to infer **B** from **A** iff

a. **A** is compatible with S's factual beliefs and $\mathbf{A} \supset \mathbf{B}$ is entailed by them, or
b. **A** is incompatible with S's factual beliefs and $\mathbf{A} \subseteq \mathbf{B}$.

This principle can be codified by a refinement operation that lifts a belief state **S** into refined belief state $\Uparrow\mathbf{S}$ as the set of all sequences where every world in **S** comes before any world not in **S**:

Refinement
$\Uparrow\mathbf{S} = \{\sigma : \forall w \in \mathbf{S} : \forall w' \notin \mathbf{S} : w >_\sigma w'\}$

Refining a coarse state maps it to a refined state that encodes the original's factual information as well as the inferential dispositions it determines. Let's look at an example. Suppose we have the following model:



Now, the refinement of **S** consists of all sequences in which w_1, w_2 come before w_3, w_4. Thus:

$$\Uparrow\mathbf{S} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4 \rangle & \langle w_1, w_2, w_4, w_3 \rangle \\ \langle w_2, w_1, w_3, w_4 \rangle & \langle w_2, w_1, w_4, w_3 \rangle \end{Bmatrix}$$

We can prove that $\Uparrow\mathbf{S}$ encodes the inferential dispositions that distinguish **S**-possible worlds from **S**-ruled out worlds, and no inferential dispositions distinguishing between **S**-possible worlds or between **S**-ruled out worlds:

Fact 1 (Refinement and Inferential Dispositions)

$\Uparrow\mathbf{S}$ encodes all the inferential dispositions determined by **S**.

Proof. Let $\mathbf{A} \mapsto \mathbf{B}$ be an arbitrary inferential disposition determined by **S**. Either $\mathbf{SA} = \varnothing$ or $\mathbf{SA} \neq \varnothing$. If the former, then $\mathbf{A} \subseteq \mathbf{B}$, and thus, for any $\sigma : \sigma(\mathbf{A}) \in \mathbf{B}$. Thus, trivially, $\Uparrow\mathbf{S}$ encodes $\mathbf{A} \mapsto \mathbf{B}$. If the latter, then $\mathbf{SA} \subseteq \mathbf{B}$. But then it follows that $\forall\sigma \in \Uparrow \mathbf{S} : \sigma(\mathbf{A}) \in \mathbf{B}$. And that is sufficient for $\Uparrow\mathbf{S}$ to encode $\mathbf{A} \mapsto \mathbf{B}$.

Let's consider an example of a refined content that encodes some inferential dispositions not determined by its factual content:

$$\mathtt{X} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4 \rangle \\ \langle w_2, w_1, w_3, w_4 \rangle \end{Bmatrix}$$

Notice, first of all, that the factual content of $\mathtt{X}$ is $\{w_1, w_2\}$—the set of worlds that appear first in any of its sequences. Thus, by Factual Determination and our refinement operation, it should only encode those inferential dispositions that distinguish between worlds in $\{w_1, w_2\}$ and those not in that set. But $\mathtt{X}$ encodes a disposition that distinguishes between worlds outside of its factual content, namely:

$$\{w_3, w_4\} \mapsto \{w_3\}$$

Thus, such a refined content represents the inferential dispositions of no possible agent. This means that such a refined content cannot be fully believed, but it can be left open (more on belief and leaving open below).

Together, the pair of operations Refinement and Coarsening (which, recall, delivers the factual content encoded by a set of sequences) is idempotent, which is sufficient to establish that there is no new information contained in the refinement of a coarse state.

Coarsening
$\Downarrow \mathtt{A} = \{w : \exists\sigma \in \mathtt{A} : \sigma_w = w\}$

Idempotence
$\Downarrow\Uparrow\mathbf{A} = \mathbf{A}$

We can also see that the only refined belief states that can model the inferential dispositions of an agent are stable in the following sense:

Stability
A refined state $\mathtt{A}$ is stable iff $\Uparrow\Downarrow\mathtt{A} = \mathtt{A}$.

Stability will be a useful property for understanding why this theory predicts violations of Conditionalization^{-} (a topic I return to below). With these results in hand, we turn next to refined belief.

4.4 Refined Belief

Recall our notions of refined belief and leaving open, from Chapter 3:

Belief (Refined)
S believes the refined content $\mathtt{A}$ iff S has every inferential disposition $\mathtt{A}$ encodes.

Leaving Open (Refined)
S leaves open $\mathtt{A}$ iff S lacks some inferential disposition $\bar{\mathtt{A}}$ encodes.

We can now prove that these notions are equivalent to the standard Hintikkan (Hintikka 1961) model-theoretic notions of belief and its dual (leaving open), where **S** is S's factual belief state:

Belief as Entailment (Refined)
S believes some refined content $\mathtt{A}$ iff $\Uparrow\mathbf{S} \subseteq \mathtt{A}$.

Leaving Open as Compatibility (Refined)
S leaves open $\mathtt{A}$ iff $\Uparrow\mathbf{S} \cap \mathtt{A} \neq \varnothing$.

We do so by proving the following intermediary results:

Fact 2 (Necessity)

$\Uparrow\mathbf{S} \subseteq \mathtt{Y}$ iff for each $w \in \mathbf{S}$, if $\mathbf{A} \mapsto_w \mathbf{B}$ is encoded by $\mathtt{Y}$, then it is among S's inferential dispositions.

Proof left-to-right. Suppose $\Uparrow\mathbf{S} \subseteq \mathtt{Y}$ and that there is a $w \in \mathbf{S}$, such that there is a non-trivial $\mathbf{A} \mapsto_w \mathbf{B}$ encoded by $\mathtt{Y}$ that is not among S's inferential dispositions. It follows that $\Uparrow \mathbf{S} \nsubseteq \mathbf{A} \mapsto_w \mathbf{B}$. So there is a $\sigma \in \Uparrow\mathbf{S}$ whose first world is w and is such that $\sigma(\mathbf{A}) \notin \mathbf{B}$. But since $\Uparrow \mathbf{S} \subseteq \mathtt{Y}$, it follows that $\sigma \in \mathtt{Y}$ as well. But since $\mathbf{A} \mapsto_w \mathbf{B}$ is encoded by $\mathtt{Y}$, it follows that $\mathtt{Y} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$, in which case $\sigma \notin \mathtt{Y}$. Contradiction.

Proof right-to-left. Suppose for each $w \in \mathbf{S}$, if $\mathbf{A} \mapsto_w \mathbf{B}$ is encoded by $\mathtt{Y}$, then it is among S's inferential dispositions, and that $\Uparrow\mathbf{S} \nsubseteq \mathtt{Y}$. We assume that $\mathbf{S}$ is non-empty and that $\mathtt{Y}$ encodes some non-trivial dispositions. From $\Uparrow\mathbf{S} \nsubseteq \mathtt{Y}$, it follows that there is a $\sigma \in \Uparrow\mathbf{S}$ that is not in $\mathtt{Y}$. Let w be its first world and suppose its first $\mathbf{A}$-world is a $\bar{\mathbf{B}}$-world. Then, $\Uparrow\mathbf{S} \nsubseteq \mathbf{A} \mapsto_w \mathbf{B}$. But, since $\sigma \notin \mathtt{Y}$, it may still be the case that $\mathtt{Y} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$. Contradiction.

Fact 3 (Possibility)

$\Uparrow\mathbf{S} \cap \mathtt{Y} \neq \varnothing$ iff for some $w \in \mathbf{S}$, there is an $\mathbf{A} \mapsto_w \mathbf{B}$ encoded by $\bar{\mathtt{Y}}$ that is not among S's inferential dispositions.

Proof left-to-right. Suppose $\Uparrow\mathbf{S} \cap \mathtt{Y} \neq \varnothing$ and that for all $w \in \mathbf{S}$, if $\mathbf{A} \mapsto_w \mathbf{B}$ is an inferential disposition encoded by $\bar{\mathtt{Y}}$, then it is among S's inferential dispositions. Since $\Uparrow\mathbf{S} \cap \mathtt{Y} \neq \varnothing$, we know there is at least one σ in their intersection; call it σ^* and suppose that its first world is w, $w \in \mathbf{S}$, and that $\sigma^*(\mathbf{A}) \notin \mathbf{B}$. It follows that $\Uparrow\mathbf{S} \nsubseteq \mathbf{A} \mapsto_w \mathbf{B}$, and thus that $\mathbf{A} \mapsto_w \mathbf{B}$ is not among S's inferential dispositions. Still, since $\bar{\text{Y}} \cap \text{Y} = \varnothing$, it may be that $\bar{\text{Y}} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$, in which case it is an inferential disposition encoded by $\bar{\text{Y}}$. But then since $w \in \mathbf{S}$, by our second assumption above, it follows that $\mathbf{A} \mapsto_w \mathbf{B}$ is among S's inferential dispositions. Contradiction.

Proof right-to-left. Suppose for some $w \in \mathbf{S}$, there is some disposition $\mathbf{A} \mapsto_w \mathbf{B}$ encoded by $\bar{\text{Y}}$ that is not among S's inferential dispositions and $\Uparrow\mathbf{X} \cap \text{Y} = \varnothing$. Thus, $\bar{\text{Y}} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$ and $\Uparrow\mathbf{S} \nsubseteq \mathbf{A} \mapsto_w \mathbf{B}$. But since $\Uparrow\mathbf{X} \cap \text{Y} = \varnothing$, every $\sigma \in \Uparrow\mathbf{S}$ must be in $\bar{\text{Y}}$. And thus since $\bar{\text{Y}} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$, it follows that $\Uparrow\mathbf{S} \subseteq \mathbf{A} \mapsto_w \mathbf{B}$. Contradiction.

We are now in a position to see how the theory generates results regarding non-factual ignorance, subjunctive bounding, and indicative bounding.

Summary: *Refined belief*

We can define a formal refinement operator that lifts factual belief states (sets of worlds) to refined belief states (sets of sequences) that encode the inferential dispositions determined by the former. Having done this, we can give a Hintikka model of belief for refined contents.

4.5 Applications

4.5.1 Non-Factual Ignorance

Recall the case from Chapter 3, where we have a fair, unflipped coin. In such a case, it seems we leave open the possibility of both:

(10) If the coin had been flipped, it would have landed heads.

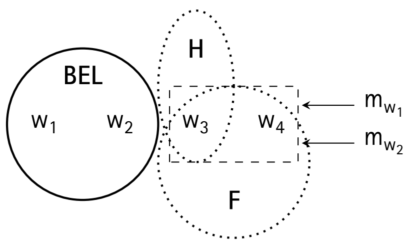
$$F \Rightarrow_m H$$

(11) If the coin had been flipped, it would not have landed heads.

$$F \Rightarrow_m \neg H$$

And furthermore, in such a case, since we know the coin is fair, there is nothing we could learn that would ensure that we believe either conditional—both conditionals are unbelievable. This is a case of non-factual ignorance, since in this case we know both conditionals are non-factual (which explains why we cannot come to believe either) and yet still leave open the possibility of both. In Chapter 3, I showed how Inferential Dispositionalism predicts these observations. Here, I show how to model this result within our sequence semantic theory of conditionals.

Let's embed this within a simplified model. Suppose that we have four worlds, w_1, w_2, w_3, w_4, and that your belief state is $\mathbf{BEL} = \{w_1, w_2\}$. Let m be the domain function of the conditional, and suppose $m_{w_1} = m_{w_2} = \{w_3, w_4\}$. Finally, suppose that **F** (that the coin is flipped) is true at w_3, w_4 and **H** (that the coin lands heads) is true at w_3:



It follows that:

$$\Uparrow\mathbf{BEL} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4\rangle & \langle w_2, w_1, w_3, w_4\rangle \\ \langle w_1, w_2, w_4, w_3\rangle & \langle w_2, w_1, w_4, w_3\rangle \end{Bmatrix}$$

This state encodes the inferential dispositions:

- $\{w_1, w_2, w_3, w_4\} \mapsto \{w_1, w_2\}$
- $\{w_1, w_3, w_4\} \mapsto \{w_1\}$
- $\{w_2, w_3, w_4\} \mapsto \{w_2\}$

Focusing just on the part of $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$ that is compatible with $\Uparrow\mathbf{BEL}$, we have:

$$\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H} = \begin{Bmatrix} \ldots & \\ \langle w_1, w_2, w_3, w_4 \rangle & \langle w_2, w_1, w_3, w_4 \rangle \\ & \ldots \end{Bmatrix}$$

This in addition encodes the inferential disposition:

$$\{w_3, w_4\} \mapsto \{w_3\}$$

We can see that $\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H} \neq \varnothing$, and thus, by Leaving Open as Compatibility (Refined), S leaves open $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$. Furthermore, notice that your inferential dispositions are compatible with those encoded by $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$, since you could coherently add the disposition to infer $\{w_3\}$ from $\{w_3, w_4\}$ to your inferential dispositions.

However, you could not come to believe $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$ without giving up something you already believe. To see why, notice that nothing you could factually learn would generate a set of inferential dispositions that included this one.

- Suppose you rule out w_2. Then, your refined belief state would look like this:

$$\Uparrow\mathbf{BEL} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4 \rangle \\ \langle w_1, w_2, w_4, w_3 \rangle \\ \langle w_1, w_3, w_2, w_4 \rangle \\ \langle w_1, w_3, w_4, w_3 \rangle \\ \langle w_1, w_4, w_2, w_3 \rangle \\ \langle w_1, w_4, w_3, w_2 \rangle \end{Bmatrix}$$

  However, notice that this state does not encode the disposition to infer $\{w_3\}$ from $\{w_3, w_4\}$ since it contains sequences where w_4 comes before w_3. Again, the reason is that such a disposition would distinguish between worlds (w_3, w_4) that you have ruled out.
- The same reasoning applies equally to ruling out w_1 rather than w_2.

Thus, our model theory illuminates the notion of non-factual ignorance.

4.5.2 Subjunctive Bounding

We saw in Chapter 3 that our theory avoids the subjunctive bounding puzzle by invalidating Conditionalization^{-}. With our model theory in hand, we can now say more precisely why this is the case.

Consider the case in which you know Smith didn't flip the coin, but you aren't sure whether the coin is fair, double-headed, or double-tailed. In such a case, could you come to believe (12)?

(12) If Smith had flipped the coin, it would have landed heads.

$$F \Rightarrow H$$

Suppose that $\mathbf{W} = \{w_1, w_2, w_3, w_4, w_5\}$, and your belief state $\mathbf{BEL} = \{w_1, w_2, w_3\}$. Suppose that the coin is fair at w_1, double-tailed at w_2, and double-headed at w_3. Finally, suppose **F** (the proposition that Smith flipped the coin) is true at w_4, w_5, and **H**, the proposition that the coin landed heads, is true at w_4. Then,

- $m_{w_1} = \{w_4, w_5\}$ Since the coin is fair at w_1
- $m_{w_2} = \{w_5\}$ Since the coin is double-tailed at w_2
- $m_{w_3} = \{w_4\}$ Since the coin is double-headed at w_3

Next, note that your refined belief state (the state encoding your inferential dispositions) is:

$$\Uparrow\mathbf{BEL} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4, w_5 \rangle & \langle w_2, w_1, w_3, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_4, w_5 \rangle \\ \langle w_1, w_2, w_3, w_5, w_4 \rangle & \langle w_2, w_1, w_3, w_5, w_4 \rangle & \langle w_3, w_1, w_2, w_5, w_4 \rangle \\ \langle w_1, w_3, w_2, w_4, w_5 \rangle & \langle w_2, w_3, w_1, w_4, w_5 \rangle & \langle w_3, w_2, w_1, w_4, w_5 \rangle \\ \langle w_1, w_3, w_2, w_5, w_4 \rangle & \langle w_2, w_3, w_1, w_5, w_4 \rangle & \langle w_3, w_2, w_1, w_5, w_4 \rangle \end{Bmatrix}$$

Remember that:

$$\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H} = \{\sigma : \sigma(m_{\sigma}^{\mathbf{A}}) \in \mathbf{B}\}$$

Now look at the relevant subset of $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$ compatible with $\Uparrow\mathbf{BEL}$:

$$\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H} = \left\{\begin{array}{ll} \ldots & \\ \langle w_1, w_2, w_3, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_4, w_5 \rangle \\ \langle w_1, w_3, w_2, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_5, w_4 \rangle \\ & \langle w_3, w_2, w_1, w_4, w_5 \rangle \\ & \langle w_3, w_2, w_1, w_5, w_4 \rangle \\ & \ldots \end{array}\right\}$$

We can see here that your refined belief state is compatible with, but does not entail, $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$. Thus, we predict (correctly) that you leave open the possibility that if the coin had been flipped, it would have landed heads, but you do not believe this.

What does it take to believe this? Well, the theory says you must come to be inferentially disposed in accordance with it. That means for each world in your belief state, you must have the dispositions encoded by the conditional at that world. Let's suppose that we could model belief update by conditionalization. This means that we could model your posterior belief state, after coming to believe $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$, as the intersection $\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$. The result would be:

$$\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H} = \left\{\begin{array}{ll} \langle w_1, w_2, w_3, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_4, w_5 \rangle \\ \langle w_1, w_3, w_2, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_5, w_4 \rangle \\ & \langle w_3, w_2, w_1, w_4, w_5 \rangle \\ & \langle w_3, w_2, w_1, w_5, w_4 \rangle \end{array}\right\}$$

However, notice that the resulting refined belief state does not encode the inferential dispositions determined by any factual belief state. Therefore, it is not a genuine refined belief state! (Remember, refined belief states are just models of inferential dispositional states.) In particular, the offending sequences are:

$$\begin{array}{l} \langle w_1, w_2, w_3, w_4, w_5 \rangle \\ \langle w_1, w_3, w_2, w_4, w_5 \rangle \end{array}$$

This pair encodes the disposition: $\mathbf{F} \mapsto_{w_1} \mathbf{H}$ (to infer that the coin landed heads from that Smith flipped it at worlds like w_1). But this disposition is not grounded in any factual belief. Indeed, it distinguishes between worlds that the corresponding factual belief state ($\{w_1, w_3\}$) rules out, ranking $\mathbf{FH}$-worlds above $\mathbf{F\bar{H}}$-worlds.

We have motivated the thought that belief update cannot proceed by conditionalization. But under what conditions, then, do we come to believe non-factual contents? As I'll now argue, belief updating must violate Conditionalization^{-}.

Consider the conditions under which you would believe $\texttt{F} \Rightarrow_{\mathrm{m}} \texttt{H}$. You would need to be in a belief state $\mathbf{BEL}^{*}$ that determines a set of inferential dispositions that include all those encoded by the conditional at each world in $\mathbf{BEL}^{*}$. We can now see again why leaving w_1 in your belief state will not work. If you come to believe that the coin is either fair- or double-headed, your belief state will be $\mathbf{BEL}^{*} = \{w_1, w_3\}$; and we will have:

$$\Uparrow\mathbf{BEL}^{*} = \begin{Bmatrix} \langle w_1, w_2, w_3, w_4, w_5 \rangle & \langle w_3, w_1, w_2, w_4, w_5 \rangle \\ \langle w_1, w_2, w_3, w_5, w_4 \rangle & \langle w_3, w_1, w_2, w_5, w_4 \rangle \\ \langle w_1, w_3, w_2, w_4, w_5 \rangle & \langle w_3, w_2, w_1, w_4, w_5 \rangle \\ \langle w_1, w_3, w_2, w_5, w_4 \rangle & \langle w_3, w_2, w_1, w_5, w_4 \rangle \end{Bmatrix}$$

But we can see that $\Uparrow\mathbf{BEL}^{*} \not\subseteq \texttt{F} \Rightarrow_{\mathrm{m}} \texttt{H}$. Thus, what we need is to also eliminate w_1 from **BEL**, which amounts to coming to believe that the coin is double-headed. This yields the state $\mathbf{BEL}^{+} = \{w_3\}$ whose refinement does entail $\texttt{F} \Rightarrow_{\mathrm{m}} \texttt{H}$:

$$\Uparrow\mathbf{BEL}^{+} = \begin{Bmatrix} \langle w_3, w_1, w_2, w_4, w_5 \rangle \\ \langle w_3, w_1, w_2, w_5, w_4 \rangle \\ \langle w_3, w_2, w_1, w_4, w_5 \rangle \\ \langle w_3, w_2, w_1, w_5, w_4 \rangle \end{Bmatrix}$$

So, in order to believe that if the coin had been flipped, it would have landed heads, you have to come to believe that the coin was double-headed. But since this conditional is weaker than the claim that the coin was double-headed (since you can leave open the possibility of the conditional even after you've ruled out that the coin is double-headed) and thus we predict failures of Conditionalization^{-}.

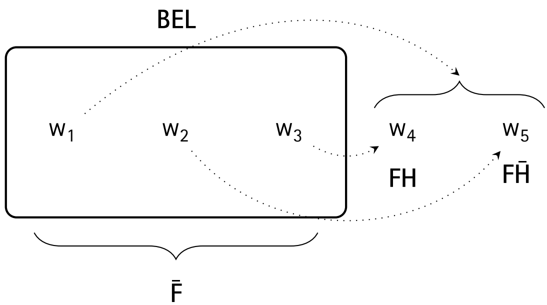
If we do not come to learn some refined content by intersecting it with our refined belief state, what formal operation will model coming to believe refined contents? I propose the following:[9]

[9] This yields a result similar to that of Kaufmann 2015's notion of "deep conditioning"—see Kaufmann 2015: 86–88.

Belief Update

S comes to believe A iff S's belief state changes to $\mathbf{BEL}_S \cap \mathbf{X}$, where $\mathbf{X}$ is the weakest proposition such that $\Uparrow(\mathbf{BEL}_S \cap \mathbf{X}) \neq \varnothing$ entails A.

Let's look at how this predicts the case above. Remember, we start out with the following belief state, where w_4 is a world where the coin is flipped and lands heads, and w_5 is a world where the coin is flipped and lands tails. And, remember, the relevant domain of flip-worlds at w_1 is $\{w_4, w_5\}$ (since the coin is fair there), and at w_2 is $\{w_5\}$ (since the coin is double-tailed there), and at w_3 is $\{w_4\}$ (since the coin is double-headed there). Letting the dotted arrow indicate m^{F}-worlds relative to each of our starting worlds in **BEL**, we have:



To learn $\mathtt{F} \Rightarrow_{\mathtt{m}} \mathtt{H}$ is to become inferentially disposed to infer **H** from its domain given **F**. And this just is to learn the minimal factual proposition that would ensure you have such inferential dispositions. And here, as we saw above, the minimal factual information you could learn that would ensure you're disposed to infer **H** from its domain given **F** is that the coin is double-headed—that is, to eliminate both worlds w_1 and w_2 from your belief state.

Given Belief Update, we predict Strong Sufficiency and violations of Conditionalization$^-$, the latter of which we have already seen.[10]

[10] See also Goldstein & Santorio 2021 for a similar approach that rejects Conditionalization$^-$ that extends the theory to epistemic modals.

Fact 4 (Belief Update Predicts Strong Sufficiency)

If $\mathbf{BEL} \subseteq \bar{\mathbf{A}}$ and $\Uparrow\mathbf{BEL} \nsubseteq \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$, then the weakest $\mathbf{X}$ such that $\Uparrow(\mathbf{BEL} \cap \mathbf{X}) \subseteq \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is $\mathbf{A} \mathbin{\Box\!\!\rightarrow}_{\mathbf{m}} \mathbf{B}$.

Define $\mathbf{A} \mathbin{\Box\!\!\rightarrow}_{\mathbf{m}} \mathbf{B} = \{w : m_w^{\mathbf{A}} \subseteq \mathbf{B}\}$.

Proof. First, we establish that $\Uparrow(\mathbf{BEL} \cap \mathbf{A} \mathbin{\Box\!\!\rightarrow}_{\mathbf{m}} \mathbf{B}) \subseteq \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$. $\mathbf{BEL} \cap \mathbf{A} \mathbin{\Box\!\!\rightarrow}_{\mathbf{m}} \mathbf{B}$ contains only worlds such that $m_w^{\mathbf{A}} \subseteq \mathbf{B}$. But then every sequence σ starting with some such world will be such that $\sigma(m_\sigma \cap \mathbf{A}) \in \mathbf{B}$.

Next, we establish that there is no $\mathbf{X}$ such that $\mathbf{BEL} \cap \mathbf{X} \nsubseteq \mathbf{BEL} \cap \mathbf{A} \mathbin{\Box\!\!\rightarrow}_{\mathbf{m}} \mathbf{B}$ for which $\Uparrow(\mathbf{BEL} \cap \mathbf{X}) \subseteq \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$. Assume otherwise. Then $\mathbf{BEL} \cap \mathbf{X}$ contains some worlds w at which $m_w^{\mathbf{A}} \nsubseteq \mathbf{B}$. Since, by assumption, $\mathbf{BEL} \subseteq \bar{\mathbf{A}}$, we know that $w \notin \mathbf{A}$. So, there must be a $\sigma \in \Uparrow(\mathbf{BEL} \cap \mathbf{X}) : \sigma(m_\sigma \cap \mathbf{A}) \notin \mathbf{B}$. But then $\Uparrow(\mathbf{BEL} \cap \mathbf{X}) \nsubseteq \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$.

4.5.3 Indicative Bounding

Finally, we are now in a position to illustrate how we avoid the indicative bounding puzzle:

Weak Sufficiency*

Coming to accept $A \rightarrow_m B$ (when you do not rule out **AB**) and m is bounded by your beliefs is to come to believe nothing more than that $A \supset B$ is true.

Strength

It is possible that $A \supset B$ is true and $A \rightarrow B$ is false.

Conditionalization$^+$

Coming to believe A is at least to rule out all possibilities in which A is false.

As discussed in Chapter 3, remember that the reason we avoid the contradiction is that it does not follow from Strength that there are possible worlds in which $\mathbf{A} \supset \mathbf{B}$ is true and $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is false. Hence, if we measure information gain in terms of which worlds you rule out, Weak Sufficiency* will be true. By contrast, it does follow from Strength that there are possible sequences in which $\mathbf{A} \supset \mathbf{B}$ is true and $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is false. But if we measure information gain in terms of which sequences you rule out, Weak Sufficiency* will be false.

We can now articulate this point more precisely, using our notion of factual truth defined above:

$$\mathtt{A} \text{ is } \begin{cases} \text{factually true at } w & \text{if } \forall\sigma : \sigma_w = w : \sigma \in \mathtt{A} \\ \text{factually false at } w & \text{if } \forall\sigma : \sigma_w = w : \sigma \notin \mathtt{A} \\ \text{neither at } w & \text{otherwise} \end{cases}$$

We can now distinguish two notions of logical strength:

> *A* is factually stronger than *B* iff there are worlds *w* at which *A* is factually true and *B* is factually false.
>
> *A* is refined stronger than *B* iff there are sequences σ at which *A* is true and *B* is false.

And, we can correspondingly distinguish two senses of Weak Sufficiency*:

> **Factual Weak Sufficiency***
>
> Coming to accept $A \rightarrow_m B$ (when you do not rule out **AB**) and *m* is bounded by your beliefs is to come to believe nothing factually stronger than $\mathbf{A} \supset \mathbf{B}$.
>
> **Refined Weak Sufficiency***
>
> Coming to accept $A \rightarrow_m B$ (when you do not rule out **AB**) and *m* is bounded by your beliefs is to come to believe nothing refined stronger than $\mathbf{A} \supset \mathbf{B}$.

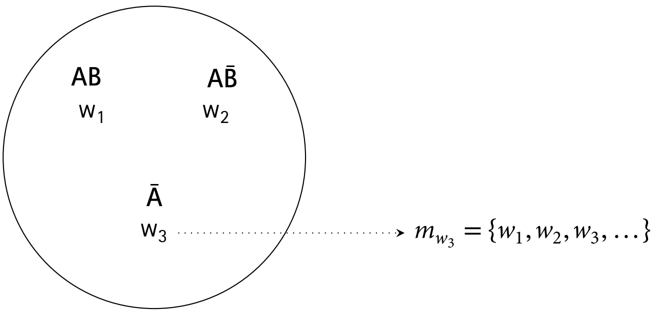
We predict that Factual Weak Sufficiency* is true, but Refined Weak Sufficiency* is false. Focus on Factual Weak Sufficiency* first. To see why this holds, we must remember two things. The first is the bounding condition:

m is bounded by **BEL** iff $\forall w \in \mathbf{BEL} : \mathbf{BEL} \subseteq m_w$

The second is the Factual Determination condition, which, recall, is equivalent to the view that $\Uparrow$**BEL** gives the refined belief state that encodes the inferential dispositions determined by **BEL**:

Refinement
$\Uparrow\mathbf{BEL} = \{\sigma : \forall w \in \mathbf{BEL} : \forall w' \notin \mathbf{BEL} : w >_\sigma w'\}$

The crucial point, again, is that $\Uparrow$**BEL** comprises every sequence in which worlds in **BEL** appear before worlds not in **BEL**. Let's now look at a simple example to illustrate Factual Weak Sufficiency*:



Note that, since m is bounded by **BEL** (by assumption), m_{w_3} contains at least w_1, w_2, w_3. To begin, notice that $\Uparrow$**BEL** does not entail $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$—it is factually false at w_2 (by Modus Ponens). Given Belief Update (reprinted here), to come to believe $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is to believe the weakest proposition **X** such that $\Uparrow(\mathbf{BEL} \cap X)$ entails $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$.

Belief Update

S comes to believe $\mathtt{A}$ iff S's belief state changes to $\mathbf{BEL}_S \cap X$, where **X** is the weakest proposition such that $\Uparrow(\mathbf{BEL}_S \cap \mathbf{X}) \neq \varnothing$ entails $\mathtt{A}$.

This proposition must rule out at the very least w_2, where the conditional is factually false. And we need not rule out w_1, where the conditional is factually true (by Strong Centering). So, we're left with

w_3. But now, we can show that once we've removed w_2 from **BEL**, its refinement already entails $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$!

To see why, consider:

$$\mathbf{BEL}^* = \{w_1, w_3\}$$

$$\Uparrow\mathbf{BEL}^* = \begin{Bmatrix} \langle w_1, w_3, \ldots \rangle \\ \langle w_3, w_1, \ldots \rangle \end{Bmatrix}$$

Now, we just have to see whether the first $m_{w_1}^{\mathbf{A}}$-world in any w_1-led sequence is a **B**-world, and whether the first $m_{w_3}^{\mathbf{A}}$-world in any w_3-led sequence is a **B**-world. And both hold:

First $m_{w_1}^{\mathbf{A}}$-world is a **B**-world

$$\Uparrow\mathbf{BEL}^* = \begin{Bmatrix} \langle w_1, w_3, \ldots \rangle \\ \langle w_3, w_1, \ldots \rangle \end{Bmatrix}$$

First $m_{w_3}^{\mathbf{A}}$-world is a **B**-world

Thus, the weakest factual proposition you need to learn to come to believe $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is one that is false only at $\mathbf{A\bar{B}}$-worlds, and this is the material conditional $\mathbf{A} \supset \mathbf{B}$!

Here is the proof of this result.

Fact 5 (Bounded Worldly Falsity)
If $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$ and m is bounded by **BEL**, then for any $w \in \mathbf{BEL}$ such that $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is false at w, $\mathbf{A} \supset \mathbf{B}$ is false at w.

Suppose $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$ and m is bounded by **BEL**. Let $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ be false at w, for some $w \in \mathbf{BEL}$. It follows that $\forall \sigma : \sigma_w = w : \sigma \notin \mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$. There are two possibilities: (i) that $m_w^{\mathbf{A}} \subseteq \bar{\mathbf{B}}$ or (ii) that $w \in \mathbf{A\bar{B}}$. The first possibility is incompatible with our first assumption; since $\mathbf{BEL} \subseteq m_w$ and $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$, it follows that $m_w^{\mathbf{A}} \nsubseteq \mathbf{B}$. So, all such worlds are $\mathbf{A\bar{B}}$-worlds, which are worlds at which $\mathbf{A} \supset \mathbf{B}$ is false.

Then, given Belief Update, it follows that the propositional information one gains when one learns an indicative conditional just is the truth of the corresponding material conditional.

Fact 6 (Propositional Weak Sufficiency*)
If $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$, $\Uparrow\mathbf{BEL} \nsubseteq \mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$, and m is bounded by **BEL**, then the weakest **X** such that $\Uparrow(\mathbf{BEL} \cap \mathbf{X}) \subseteq \mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ is $\mathbf{A} \supset \mathbf{B}$.

However, notice that if we instead measure information gain in terms of refined content, it follows from Strength that you will learn more than just that $\mathtt{A} \supset \mathtt{B}$ is true.

Fact 7 (Refined Weak Sufficiency* Fails)
It may be that $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$, $\Uparrow\mathbf{BEL} \nsubseteq \mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$, and m is bounded by **BEL** and there is a $\sigma \in \Uparrow\mathbf{BEL}$ at which $\mathtt{A} \supset \mathtt{B}$ is true that is not in $\Uparrow \mathbf{BEL}^*$, your belief state after coming to believe $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$.

> Suppose $\mathbf{BEL} \cap \mathbf{AB} \neq \varnothing$, $\Uparrow\mathbf{BEL} \nsubseteq \mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$, and m is bounded by **BEL**. As long as some $w \in \mathbf{BEL}$ are such that $m_w^{\mathbf{A}} \nsubseteq \mathbf{B}$, then, given Strength, there will be a sequence $\sigma \in \Uparrow\mathbf{BEL}$ at which $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ is false and at which $\mathtt{A} \supset \mathtt{B}$ is true. This is a $\bar{\mathbf{A}}$-led sequence whose first $m_w^{\mathbf{A}}$-world is a $\bar{\mathbf{B}}$-world. However, by hypothesis, $\Uparrow\mathbf{BEL}^*$ contains no such sequences, since $\Uparrow\mathbf{BEL}^* \subseteq \mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$. Therefore, $\Uparrow\mathbf{BEL}^*$ rules out some sequence that $\Uparrow\mathbf{BEL}$ previously didn't rule out at which $\mathtt{A} \supset \mathtt{B}$ is true.

Again, we can either measure information gain in terms of propositional content (my preferred approach) or refined content. Measured in terms of propositional content, each of Factual Weak Sufficiency*, Strength, and Conditionalization$^+$ is true, and we avoid the bounding puzzle because it does not follow from Strength that there is a possible world at which $\mathbf{A} \supset \mathbf{B}$ is true and $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ is factually false, given the antecedent of Factual Weak Sufficiency*. Measuring information gain in terms of refined content, it does follow from Strength that there is a possibility (a sequence of worlds) at which $\mathbf{A} \supset \mathbf{B}$ is true and $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ is false, in which case Refined Weak Sufficiency* is falsified. But this wasn't the sense of Weak Sufficiency* that mattered intuitively.

4.6 Summary and Next Steps

Here is a statement of our formal semantics for conditionals, so far:

Baseline Semantics
$[\![A \to B]\!]^{m,\sigma} = 1$ iff $\sigma(m^{\mathbf{A}}_{\sigma}) \in \mathbf{B}$

This will need to be refined in a few ways, as we'll see in the coming chapters. But as a preview, let me explain how we got here and where we're going.

The shift to defining the truth conditions of conditionals relative to sequences rather than worlds offered us several important advantages. One was that we could model the proposal that conditionals encode inferential dispositions; the second was that we could model violations of Conditionalization^{-} (to avoid the subjunctive bounding puzzle); and the third was that we could precisely state why the resulting theory avoids the indicative bounding puzzle (by capturing the strength of indicative conditionals compared with material conditionals as measured in their refined content).

However, refinements are coming. In Chapter 5 and 6, I will discuss the probabilities of conditionals, which will motivate relativizing the truth conditions of conditionals to a third parameter—a partition $\mathbb{Z}$ often set by a question under discussion in the context—and decoupling the modal meaning of conditionals from their domain restriction element. In particular, following Kratzer 1986, I will hold that it comes from a covert modal $\rhd$ (which is, importantly, not a necessity modal; cf. Rothschild 2013, Mandelkern 2018). This will allow us to distinguish what are sometimes called single-case conditionals (which we have focused on exclusively) from multi-case conditionals like:[11]

(13) If Jones wakes up early enough, he has coffee before class.

I adopt the strategy for multi-case conditionals on which they involve some kind of covert adverbial quantifier—perhaps a generic quantifier

[11] The terminology comes from Kadmon 1987.

over situations—in place of the covert modal (cf. Lewis 1975, Heim 1990, von Fintel 1994, Khoo 2011).

The ultimate statement of my semantics for single-case conditionals is this:

Refined Semantics
$[\![A \to \rhd B]\!]^{m,\sigma,\mathbb{Z}} = 1$ iff $[\![\rhd B]\!]^{m^{\mathbf{A}},\sigma,\mathbb{Z}} = 1$

Where $[\![\rhd B]\!]^{m,\sigma,\mathbb{Z}} = 1$ iff $\sigma(m_w \cap \mathbb{Z}) \in \mathbf{B}$

Again, I will explain and motivate each of these refinements in Chapters 5 and 6.

PART II
PROBABILITIES

5

Probabilities of Conditionals

In the last chapter, I showed how to reduce what it is to believe or leave open the possibility of a refined content—one that encodes inferential dispositions—to properties of an agent's factual belief state. However, in addition to believing and leaving open conditionals, we also think them more or less probable. Furthermore, as we saw in Chapters 1 and 2, the probability of a conditional seems to track the corresponding conditional probability of its consequent given its antecedent. Recall the case where a fair die has been rolled but the outcome remains unknown. It is very plausible that the probability that:

(1) If the die landed on an odd, it landed on a prime.

is equal to the probability that the die landed on a prime, given that it landed on an odd, which is 2/3 (since two of the three odd outcomes are prime). Cases like these suggest the following generalization about the probabilities of conditionals (so-called because Stalnaker (1970) was the first to propose it in this form):

Stalnaker's Thesis

For all sentences A, B, and probability functions P: if $P(A) > 0$,

$$P(A \rightarrow B) = P(B|A)$$

Despite its intuitive appeal, Stalnaker's Thesis must be false. At the very least, it must be restricted to indicative conditionals, as we can see with examples like:

(2) If Oswald hadn't shot Kennedy, no one else would have.

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0006

Suppose you leave open a small possibility that Oswald didn't shoot Kennedy. Since you know Kennedy was shot, your probability that no one else shot Kennedy given that Oswald didn't is 0. But you may still regard (2) as probable.[1]

So, Stalnaker's Thesis at most applies to indicative conditionals, which, for us, are those whose modal bases are epistemic and (usually) bounded by the agent's belief state. However, there is a serious problem. Almost as soon as Stalnaker proposed this thesis, Lewis 1976 proved a series of results that establish that it can hold only in trivial models that are not rich enough to model the beliefs of most ordinary agents. For instance, Lewis showed that Stalnaker's Thesis entails that, whenever a probability function P assigns positive values to $\mathbf{AB}$ and $\mathbf{A\bar{B}}$, the propositions $\mathbf{A}$ and $\mathbf{B}$ must be probabilistically independent: $P(\mathbf{B}) = P(\mathbf{B}|\mathbf{A})$. And this is clearly implausible. Knowing nothing about how a fair die landed, the probability that it landed on an even is 0.5. But the probability that it landed on an even given that it landed on an odd is 0.

Furthermore, Lewis's first four triviality results were just the beginning of the problem. Since 1976, there have been a bevy of increasingly troubling triviality results—relying on weaker and weaker versions of Stalnaker's Thesis—raising the question of whether there is any plausible way to limit its scope, while still capturing our intuitions about the probabilities of various conditionals.

In this chapter, I explore two questions:

- **Question 1**: What is the correct statement of Stalnaker's Thesis?
- **Question 2**: Can our theory predict this version of Stalnaker's Thesis?

I begin the chapter with an informal discussion of what it is to invest some probability in a refined content. But an important question remains: Can we assign probabilities to conditional contents that match our intuitions about their probabilities?

[1] We will return to discuss the probabilities of subjunctive conditionals like (2) in Chapter 10.

It is far from obvious that this is possible. We will look at triviality results that threaten the generality of Stalnaker's Thesis in two dimensions. Diachronic triviality results raise trouble for generalizing Stalnaker's Thesis across probability functions, holding fixed a particular conditional. Synchronic triviality results raise trouble for generalizing Stalnaker's Thesis across conditionals, holding fixed a particular probability function. The remainder of this chapter will discuss how to respond to such results, and whether we can salvage a plausible restricted form of Stalnaker's Thesis in light of them.

In §5.2, I discuss various diachronic triviality results (from Lewis 1976 and others), and argue that denying Conditionalization^{-}—something we have already motivated in Chapters 3 and 4—is sufficient to avoid their implausible consequences. Then, in §5.3, I argue that various synchronic triviality results require constraining the generality of Stalnaker's Thesis: it holds only for conditionals whose antecedents and consequents are not themselves conditionals (basically, ruling out extending Stalnaker's Thesis to left- and right-nested conditionals). I argue that this constraint is also independently intuitively plausible, and thus no intuitive cost. Having motivated constraints on Stalnaker's Thesis that allow us to avoid both diachronic and synchronic triviality, in §5.4, I prove a tenability result that shows we can extend the sequence semantics of Chapter 4 to generate a constrained version of Stalnaker's Thesis that predicts our ordinary intuitions about the probabilities of indicative conditionals without falling afoul of triviality results.

5.1 Preliminaries

In addition to being possible or necessary, conditionals may also be more or less probable (or likely true). Just as we abstract away from the flavor of possibility, our interest here will be the notion of probability in abstraction from any particular flavor (evidential, objective chanciness, etc). The fact that we have clear intuitions about how likely various conditionals are reason enough to dig into the matter, even before we know precisely what ordinary speakers mean when they report such judgments.

In the usual way, we will relativize the probability of a proposition to a measure P that assigns subsets of $\mathbf{W}$ values in the unit interval and satisfies the standard axioms:

- Totality: $P(\mathbf{W}) = 1$
- Finite Additivity: For any $\mathbf{A}, \mathbf{B} \in \wp(\mathbf{W})$ such that $\mathbf{A} \cap \mathbf{B} = \varnothing$: $P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$

We define the conditional probability of $\mathbf{B}$ given $\mathbf{A}$ in the standard way, using the Ratio formula:

$$P^{|\mathbf{A}}(\mathbf{B}) = P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})}, \text{ if } P(\mathbf{A}) > 0$$

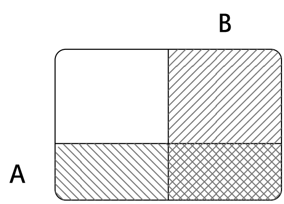
Conditional probabilities are, roughly, dispositions to update one's probabilities: the conditional probability of $\mathbf{B}$ given $\mathbf{A}$ is the probability one would assign to $\mathbf{B}$, assuming one came to assign probability 1 to $\mathbf{A}$.

Some probability functions represent the epistemic probabilities of an agent—how likely she thinks various propositions are given her evidence. Such functions are normalized on the agent's belief (or evidence) state, in the following sense:

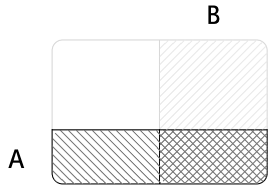
P is normalized on **BEL** iff

a. $\mathbf{X} \cap \mathbf{BEL} \neq \varnothing$ iff $P(\mathbf{X}) > 0$, and
b. $\mathbf{X} \cap \mathbf{BEL} = \varnothing$ iff $P(\mathbf{X}) = 0$.

Just as we represent belief states graphically as subsets of logical space, we can represent a probability function graphically, with the sizes of regions of logical space representing the relative weights the function assigns to various propositions:



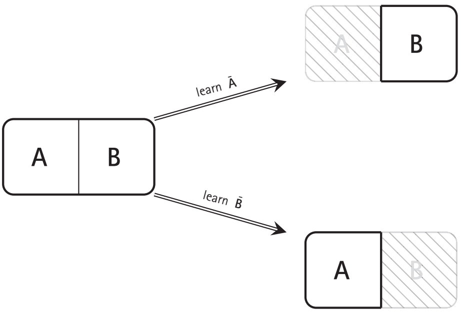
Here, this represents that **A** is about 1/3 probable and **B** is about 1/2 probable. The conditional probability of **B** given **A** is given by looking at the relative proportion of the **B** to **B̄**-space at the **A**-region (which is also 1/2):



So far, this is all old hat. What I want to do now is sketch a strategy for defining the probability of a refined content (one encoding inferential dispositions) in terms of a probability measure *P* over propositions. Our ultimate goal will be for *P* to determine a refined probability measure *Pr* that maps subsets of Ω into elements of the unit interval and obeys the standard axioms:

- Totality: $Pr(\Omega) = 1$
- Finite Additivity: For any $\mathtt{A}, \mathtt{B} \in \wp(\Omega)$ such that $\mathtt{A} \cap \mathtt{B} = \varnothing$: $Pr(\mathtt{A} \cup \mathtt{B}) = Pr(\mathtt{A}) + Pr(\mathtt{B})$

Just as ⇑**BEL** encodes the inferential dispositions determined by **BEL**, we want *Pr* to encode the probabilistic inferential dispositions determined by *P*. Just as *P* is a probability function normalized on **BEL**, *Pr* will be a probability function over sequences normalized on ⇑**BEL**. Recall, for instance, that we predict that if our belief state can be partitioned into two mutually incompatible non-empty cells, **A**/**B**, then we will be disposed to infer **B** from **Ā**, and vice versa:



Likewise, if our probability measure P is such that $P(\bar{\mathbf{A}}) > 0$, then we are disposed to assign **B** the value $P(\mathbf{B}|\bar{\mathbf{A}})$ upon coming to believe $\bar{\mathbf{A}}$. In other words, our probabilistic inferential dispositions are determined by P in accordance with the rule **Propositional Conditionalization**:

> **Propositional Conditionalization**
> An agent whose evidential probability function is P is disposed to assign **Y** probability z given **X**, where $z = P(\mathbf{Y}|\mathbf{X})$, if $P(\mathbf{X}) > 0$.

Thus, in exactly the same way that refined belief states encode inferential doxastic dispositions, a refined probability state will encode probabilistic inferential dispositions. Notice, though, that our commitment to Propositional Conditionalization does not commit us to the rationality or actuality of agents following through with these dispositions. Thus, it may be that S is disposed to assign **Y** probability $P(\mathbf{Y}|\mathbf{X})$ given **X**, but when S learns **X**, she in fact assigns **Y** some other probability. And, perhaps it is even rational to do this sometimes (contra Teller 1973; Lewis 1999; I am not here committed to conditionalization as a rational ideal).

Thomason's example from Chapter 3 helpfully illustrates this difference (see also the discussion in Chapter 3). Suppose I have full introspective access to my beliefs. I regard it as unlikely that my business partner is cheating me; since I know my beliefs, I assign probability 0 to the proposition that I believe my business partner is cheating me. Then, by Propositional Conditionalization, I am disposed to assign probability 0 to that I believe my business partner is cheating me given that my business partner is cheating me. But of course were I to learn that my business partner is cheating me, then I would assign probability 1 to that I think my business partner is cheating me. Notice that belief in the conditional goes by the former, not the latter. That is, I accept (3-a) and not (3-b):

(3) a. If my business partner is cheating me, I don't believe it.
    b. If my business partner is cheating me, I believe it.

Interestingly, the future-directed case is different; neither of these is obviously right or obviously wrong given the information provided:

(4) a. If my business partner is cheating me, I won't find out.
    b. If my business partner is cheating me, I will find out.

Given that my future beliefs are open, my acceptance of (4-a)/(4-b) will depend on my dispositions to update my beliefs about my future beliefs given certain information, and so either may be accepted or rejected, or invested with various non-trivial probabilities.

Turning back to Propositional Conditionalization, it follows that a probability measure will determine a refined probability measure, which encodes an agent's probabilistic inferential dispositions. I leave the details of how to formally define refined probability functions in terms of probability functions until §5.4. Here, I articulate the basic strategy. Start with a probability mass function p that weights worlds according to the probability that they are actual, and then extend this to a mass function over sequences that weights sequences by the probability assigned to the first world times the probability assigned to the second world given that the first world is non-actual times the probability assigned to the third world given that the first two are non-actual, and so on.[2]

Here's a simple example. Suppose $\mathbf{W} = \{w_1, w_2, w_3, w_4, w_4\}$ and $\mathbf{BEL} = \{w_1, w_2, w_3\}$. Thus,

$$\Uparrow\!\mathbf{BEL} = \left\{\begin{matrix} \langle w_1, w_2, w_3, w_4, w_5\rangle & \langle w_2, w_1, w_3, w_4, w_5\rangle & \langle w_3, w_1, w_2, w_4, w_5\rangle \\ \langle w_1, w_2, w_3, w_4, w_5\rangle & \langle w_2, w_1, w_3, w_5, w_4\rangle & \langle w_3, w_1, w_2, w_5, w_4\rangle \\ \langle w_1, w_3, w_2, w_4, w_5\rangle & \langle w_2, w_3, w_1, w_4, w_5\rangle & \langle w_3, w_2, w_1, w_4, w_5\rangle \\ \langle w_1, w_3, w_2, w_5, w_4\rangle & \langle w_2, w_3, w_1, w_5, w_4\rangle & \langle w_3, w_2, w_1, w_5, w_4\rangle \end{matrix}\right\}$$

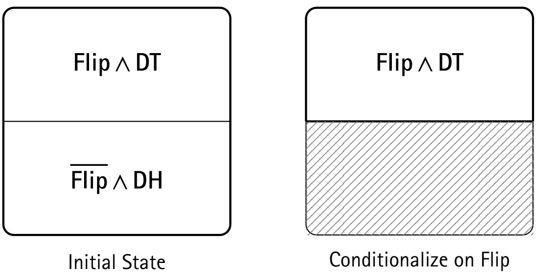
[2] This strategy is inspired by van Fraassen 1976's famous tenability result, as well as a letter from Robert Stalnaker to Richard Jeffrey quoted in full in Jeffrey 1991: 178–183. Related approaches include McGee 1989; Jeffrey 1991; Stalnaker & Jeffrey 1994; Kaufmann 2005a, 2009, 2015; Bradley 2012; Khoo & Santorio 2018; Santorio & Goldstein 2021.

Suppose $p(w_1) = 0.5, p(w_2) = 0.25, p(w_3) = 0.25, p(w_4) = p(w_5) = 0$. Then, this state encodes the disposition to assign $\{w_1\}$ probability 2/3 upon coming to believe $\{w_1, w_3\}$. This is because $P(\{w_1\}|\{w_1, w_3\}) = 2/3$. Other refined states will not obey Propositional Conditionalization in this sense, but all of the states encoding some agent's probabilistic inferential dispositions do.

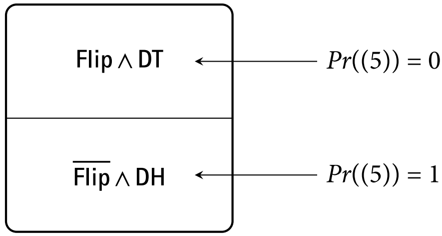
Finally, before we get to the triviality results, I want to motivate an initial constraint on Stalnaker's Thesis: it only applies to conditionals whose domains bear a certain relationship to the agent's probability function. In particular, if the agent allows that the conditional's domain may include worlds that the agent rules out, we will get violations of Stalnaker's Thesis. Here is a simple case, illustrated with a subjunctive conditional. Suppose it is 1/2 likely that Smith flipped the coin, and certain that if he flipped it, it was double-tailed, and certain that if he did not flip it, it was double-headed. Consider:

(5) If Smith had flipped the coin, it would have landed heads.

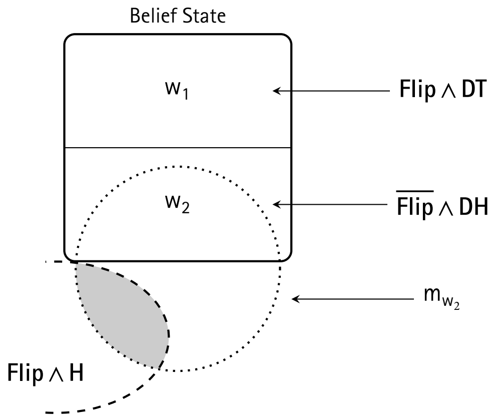
The conditional probability that the coin landed heads, given that Smith flipped it, is 0:



But the probability of (5) is intuitively greater than 0. At flip-worlds, its probability is 0, but at non-flip-worlds, its probability is 1. So, intuitively, its probability should be 1/2.



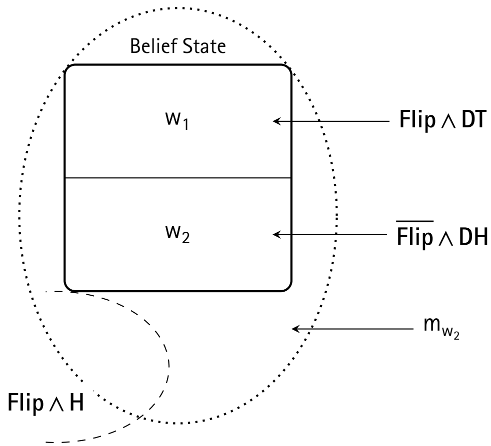
The problem here is that there are worlds compatible with our beliefs, namely $\overline{\textbf{Flip}} \wedge \textbf{DH}$-worlds, where the domain of the conditional (5) contains worlds that we have ruled out—namely, worlds where Smith flips the coin and it lands heads. Thus:



Examples like these motivate constraining Stalnaker's Thesis to conditionals whose domains are bounded by the probability function:

> m is bounded by P iff P is normalized on a state **S** such that m is bounded by **S**.
>
> m is bounded by **S** iff $\forall w \in \mathbf{S} : \mathbf{S} \subseteq m_w$.

Narrowing attention to conditionals whose modal base m is bounded by P forces m_{w_2} in the case above to contain your belief state **BEL**, and thus contain w_1:



Thus, it will follow that you fully believe $m_{w_2}^{\mathbf{Flip}} \supset \bar{\mathbf{H}}$, in which case, given the constraint of Factual Determination that your inferential dispositions distinguish worlds compatible with your belief state from worlds incompatible with it, it will follow that your probability in (5) is 0, matching your conditional probability of **H** given **Flip**. Now, this is *obviously not* the result we want to predict for (5); that's not my point here. My point is that (5) is obviously not an instance of Stalnaker's Thesis, and what makes it not an instance is that its modal base is not bounded by the relevant probability function. Thus, such examples show that the only plausible version of Stalnaker's Thesis must at the very least be constrained to conditionals whose modal bases are bounded by the relevant probability function.

In what follows, I will assume (unless otherwise noted) that, where *P* is some probability function, *Pr* is the refined probability function determined by it (that is, the measure that encodes the agent's probabilistic inferential dispositions). Given this, our starting point for Stalnaker's Thesis may be stated as follows:

Stalnaker's Thesis (Starting)

For every *Pr* and modal base *m* such that *m* is bounded by *Pr*, if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \to_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

As we'll see, even this weakened version of Stalnaker's Thesis is subject to triviality results, which will necessitate restricting its generality in various ways.

Summary: *Stalnaker's Thesis*

We outlined the formal preliminaries and the strongest plausible version of Stalnaker's Thesis. As we'll see below, even this thesis will need to be pared back in various ways.

5.2 Diachronic Triviality

Lewis 1976 proved that Stalnaker's Thesis (Starting) was implausible, since it holds only on trivial models. As we'll see, Lewis's proof also revealed the path forward for a plausible, non-trivial version of Stalnaker's Thesis; however, that lesson has, I think, been lost in the ensuing discussion. In this section, I will isolate a crucial closure assumption of Lewis's proof and show that the theory in Chapters 3 and 4 has already motivated its denial. Since Lewis's closure assumption is behind a number of stronger diachronic triviality results, denying it will be sufficient to generate a weakened version of Stalnaker's Thesis that avoids all of them.

Here, in simple form, is Lewis's triviality result. We start with Stalnaker's Thesis as stated above:

Stalnaker's Thesis (Starting)

For every Pr and modal base m such that m is bounded by Pr, if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

Next, observe that some probability functions are non-trivial in the following sense:

Non-Triviality

For some sentences A, B, probability function P,
$$P(\mathbf{AB}) > 0, P(\mathbf{A\bar{B}}) > 0, \text{ and } P(\mathbf{B}|\mathbf{A}) \neq P(\mathbf{B}).$$

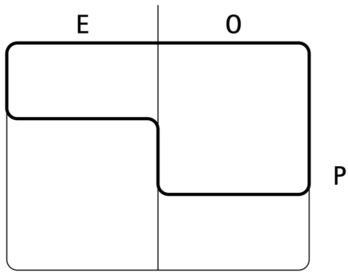
Non-Triviality is an incredibly weak observation. A fair die has been rolled but you have no evidence how it landed. It's thus possible that the die landed on an even prime, and possible it landed on an odd prime. The probability that it landed on an even given it landed on a prime is 1/3, which is lower than the probability that it landed on an even, which is 1/2:

$P(\mathbf{EP}) = 1/6$

$P(\mathbf{OP}) = 1/3$

$P(\mathbf{E}|\mathbf{P}) = 1/3$

$P(\mathbf{E}) = 1/2$



However, surprisingly, Lewis's proof shows that Stalnaker's Thesis (Starting) is incompatible with Non-Triviality. Let P be an arbitrary probability function. Assume (i) m is bounded by P, (ii) $P(\mathbf{AB}) > 0$, and (iii) $P(\mathbf{A\bar{B}}) > 0$. Throughout, Pr will be the refined probability function determined by P. Together with Stalnaker's Thesis (Starting), we can prove:

(L1) $Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = 1$

(L2) $Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\bar{\mathtt{B}}) = 0$

Given that m is bounded by P and (ii)/(iii), it follows that m is bounded by $Pr(\cdot|\mathtt{B})$ and $P(\cdot|\bar{\mathtt{B}})$. Thus, by Stalnaker's Thesis (Starting), the refined counterparts of these probability functions should assign a probability to the conditional equal to the probability of its consequent given its antecedent, establishing both (L1) and (L2) immediately:

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = Pr(\mathtt{B}|\mathtt{AB}) = 1$$

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\bar{\mathtt{B}}) = Pr(\mathtt{B}|\mathtt{A}\bar{\mathtt{B}}) = 0$$

Since *m* is bounded by *Pr*, then given Stalnaker's Thesis (Starting) and the Law of Total Probability, we have:

$$\begin{aligned} Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) &= \underbrace{Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B})}_{1} \cdot Pr(\mathtt{B}) + \underbrace{Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\bar{\mathtt{B}})}_{0} \cdot Pr(\bar{\mathtt{B}}) \\ &= Pr(\mathtt{B}) \\ &= Pr(\mathtt{B}|\mathtt{A}) \end{aligned}$$

Thus, since neither *A* nor *B* are conditionals, we establish that $P(\mathbf{B}) = P(\mathbf{B}|\mathbf{A})$, and thus since *P* was arbitrary, that Non-Triviality is false.

Lewis's triviality result shows that Stalnaker's Thesis (Starting) is too strong.[3] But this itself shouldn't trouble us—Stalnaker's Thesis is much stronger than anything we need to predict ordinary intuitions about the probabilities of conditionals. If it turned out that the only probability functions violating Stalnaker's Thesis were ones that could not model the evidential probabilities of any possible agent, then it wouldn't be much of a cost to deny it. Thus, we should distinguish Stalnaker's Thesis from:

Agential Stalnaker's Thesis

For every *Pr* that encodes the probabilistic inferential dispositions of a possible agent and modal base *m* such that *m* is bounded by *Pr*, if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

It is this constrained principle that we should care about, given our explanatory ambitions, not anything as strong as Stalnaker's Thesis (Starting). But, now, notice that Lewis's triviality result applies just as well to Agential Stalnaker's Thesis, given:

[3] An alternative response to the results is to keep Stalnaker's Thesis (Starting) and instead deny the Law of Total Probability. Trivalent theories, such as Lassiter 2020, adopt this strategy. I discuss these views in the Appendix A to this chapter.

Closure Under Conditionalization

If Pr encodes the probabilistic inferential dispositions of a possible agent, and $Pr(\mathtt{A}) > 0$, then $Pr(\cdot|\mathtt{A})$ encodes the probabilistic inferential dispositions of a possible agent.

Furthermore, it may seem that I am committed to Closure Under Conditionalization. After all, I claim that whenever P is an agent's evidential probability function, Pr encodes the probabilistic inferential dispositions determined by P. We'll see below how to generate a unique such Pr from P in the final section of this chapter, but the general idea follows the fact that factual beliefs determine inferential dispositions (the principle Factual Determination, from Chapter 3). However, I am also committed to conditionalization:

Propositional Conditionalization

An agent whose evidential probability function is P is disposed to assign $\mathbf{Y}$ probability z given $\mathbf{X}$, where $z = Pr(\mathbf{Y}|\mathbf{X})$, if $P(\mathbf{X}) > 0$.

Together, these seem to suggest Closure Under Conditionalization. But this principle, together with Agential Stalnaker's Thesis, leads right back to triviality. Given Non-Triviality, we may assume that some agent's evidential probability function P is such that $P(\mathbf{AB}) > 0$, and thus $Pr(\mathtt{AB}) > 0$. Then, by Closure Under Conditionalization, it follows that $Pr(\cdot|\mathtt{B})$ encodes the rational probabilistic inferential dispositions of a possible agent. Then, since $Pr(\mathtt{A}|\mathtt{B}) > 0$ and m is bounded by Pr, it follows from Agential Stalnaker's Thesis that:

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = Pr(\mathtt{B}|\mathtt{AB}) = 1$$

thus establishing (L1) (and similarly, (L2)).

5.2.1 Denying Closure

Lewis's result thus pits Agential Stalnaker's Thesis against Closure Under Conditionalization. Thus, we might consider weakening

Agential Stalnaker's Thesis, perhaps to quantify only over probability functions normalized on the relevant modal base. Then, we could wield the resources of contextualism to capture our intuitions without running into the diachronic triviality results (a strategy explored in Mandelkern & Khoo 2019). While I think this is a viable strategy to pursue (though see the challenges raised by Paolo Santorio in Khoo & Santorio 2018: 49–51, as well as those raised by Charlow 2019), we have already seen independent reasons for rejecting Conditionalization^{-} (see Chapter 3), and thus we can appeal to our motivations for doing so to motivate rejecting Closure Under Conditionalization. Thus, here I advocate the latter strategy: we should reject Closure Under Conditionalization.[4]

To see why, suppose as before that a fair die has been rolled and we have no evidence about how it landed. We thus think it is possible that the die landed on an even prime, and possible it landed on an odd prime. Furthermore, in accordance with Agential Stalnaker's Thesis, it seems correct that our probability that if the die landed on a prime, it landed on an even is 1/3, equal to the probability that it landed on an even given that it landed on a prime. This latter probability represents our disposition to infer that the probability that the die landed on an even is 1/3 from that it landed on a prime.

- $Pr(\mathtt{EP}) = 1/6$
- $Pr(\mathtt{OP}) = 1/3$
- $Pr(\mathtt{P} \rightarrow_{\mathtt{m}} \mathtt{E}) = Pr(\mathtt{E}|\mathtt{P}) = 1/3$

According to Closure Under Conditionalization, $Pr(\cdot|\mathtt{E})$ encodes the probabilistic inferential dispositions of a possible agent. However, we can now show that there cannot be such an agent. We show this in two steps.

First, any agent whose probabilistic inferential dispositions are encoded by $Pr(\cdot|\mathtt{E})$ must be such that their evidential probability

[4] See also Kaufmann 2015: 85–90, although whereas I see this as more evidence that refined contents are in a certain sense not real (non-factual), Kaufmann takes this as reason to work entirely within the refined space, so as not to get mismatches between the factual probabilities and refined probabilities. Also, see Goldstein & Santorio 2021, who extend this kind of theory to epistemic modals.

function P' is such that $P'(\mathbf{E}) = 1$. This is because $Pr(\cdot|\mathtt{E})$ encodes at least the disposition to infer that **E** has probability 1 from nothing. But that means that such an agent fully believes **E**, and thus fully believes $\mathbf{P} \supset \mathbf{E}$, and thus, by Factual Determination (from Chapter 3) has the disposition to infer **E** from $m_w^{\mathbf{P}}$, for every w in their belief state (remember, m is bounded by P').

Factual Determination

An agent S is disposed to infer **B** from **A** iff

a. **A** is compatible with S's factual beliefs and $\mathbf{A} \supset \mathbf{B}$ is entailed by them, or
b. **A** is incompatible with S's factual beliefs and $\mathbf{A} \vDash \mathbf{B}$.

But that is sufficient for the agent to fully believe $\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}$, and thus invest probability 1 in this conditional. So, if $Pr(\cdot|\mathtt{E})$ encodes the probabilistic inferential dispositions of some agent, $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\mathtt{E}) = 1$.

Now for the second step. We have stipulated that $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}) = Pr(\mathtt{E}|\mathtt{P}) = 1/3$. It follows, then, that $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\mathtt{E}) < 1$. To see this, suppose otherwise.

Suppose $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\mathtt{E}) = 1$. Then,

$$\begin{aligned} Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}) = \;& \underbrace{Pr(\mathtt{P} \rightarrow \mathtt{E}|\mathtt{E})}_{1} \cdot Pr(\mathtt{E}) + Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\bar{\mathtt{E}}) \cdot Pr(\bar{\mathtt{E}}) \\ = \;& 1 \cdot 1/2 + Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\bar{\mathtt{E}}) \cdot 1/2 \\ \geq \;& 1/2 > \underbrace{Pr(\mathtt{E}|\mathtt{P})}_{1/3} \end{aligned}$$

Since it cannot be the case that $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\mathtt{E}) = 1$ and $Pr(\mathtt{P} \rightarrow_{\mathrm{m}} \mathtt{E}|\mathtt{E}) < 1$, we have a counterexample to Closure Under Conditionalization. We started with the assumption that P is the evidential probability of a possible agent, and reached the conclusion that $Pr(\cdot|\mathtt{E})$ does not encode the probabilistic inferential dispositions of any agent.

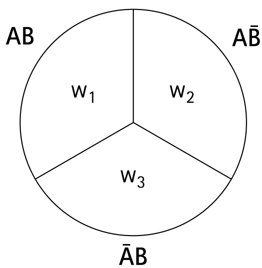
To see why this result is compatible with Propositional Conditionalization, notice that taking any agent's probability function P and conditionalizing it on any non-zero-probability proposition will result in an evidential probability function of some agent—namely, the starting agent's after she has learned that proposition. What we deny is that taking the refined probability measure that encodes that agent's probabilistic inferential dispositions and conditionalizing *it* on some non-zero-probability refined content will always result in a measure that encodes any possible agent's probabilistic inferential dispositions (again, compare Kaufmann 2015's distinction between "shallow" and "deep" conditioning):

$$\begin{tikzcd}
\text{Non-Factual Probabilities } Pr \arrow[rr, Rightarrow, "\text{conditionalization on A}"] & & Pr\,(\cdot|\,\mathbf{A}) \neq Pr' \\
\text{Factual Probabilities } P \arrow[u, "\text{determines}"] \arrow[rr, Rightarrow, "\text{conditionalization on A}"] & & P' \arrow[u, "\text{determines}"']
\end{tikzcd}$$

Where $P'(\mathbf{X}) = P(\mathbf{X}|\mathbf{A})$

Let's pause to see why we generate this result from the perspective of the model theory. Consider the following case:

> **Murder Mystery**. Dr. Jones has been found killed and we are on the job. We've narrowed the suspects to two, the **A**rtist and the **B**utler, but aren't sure whether they worked together or separately. We invest equal confidence that they worked together as that the Artist did it on his own as that the Butler did it on his own.
>
> Our belief state is **BEL** and our probability function is P. Suppose $m_{w_1} = m_{w_2} = m_{w_3} = \mathbf{BEL} = \{w_1, w_2, w_3\}$. Suppose m is bounded by P and that P invests equal confidence in each world in **BEL**.



The probability that:

(6) If the Artist did it, the Butler did, too.

$A \rightarrow_m B$

is ½. We can now show that the probability that if the Artist did it, the Butler did, too, given that the Butler did it, is less than zero: $Pr(\texttt{A} \rightarrow_{\texttt{m}} \texttt{B} | \texttt{B}) < 1$.

> $\texttt{A} \rightarrow_{\texttt{m}} \texttt{B}$ is true at some but not all *Pr*-non-zero sequences starting with w_3, since at some such sequences the first $m^{\mathbf{A}}_{w_3}$-world is a **B**-world and at some the first $m^{\mathbf{A}}_{w_3}$-world is a $\bar{\mathbf{B}}$-world. Thus, $\texttt{A} \rightarrow_{\texttt{m}} \texttt{B}$ is not true at all B-sequences that *Pr* assigns non-zero probability.
>
> $$\Uparrow \mathbf{BEL} = \begin{Bmatrix} \langle w_1, w_2, w_3 \rangle & \langle w_2, w_1, w_3 \rangle & \langle w_3, w_1, w_2 \rangle \\ \langle w_1, w_3, w_2 \rangle & \langle w_2, w_3, w_1 \rangle & \langle w_3, w_2, w_1 \rangle \end{Bmatrix}$$
>
> True at $\langle w_3, w_1, w_2 \rangle$, since its first $m^{\mathbf{A}}_{w_3}$-world is $w_1 \in \mathbf{B}$.
> False at $\langle w_3, w_2, w_1 \rangle$, since its first $m^{\mathbf{A}}_{w_3}$-world is $w_2 \notin \mathbf{B}$.

Now suppose that our agent comes to believe that the Butler did it, thus updating her probability function to P', where for all **X**, $P'(\mathbf{X}) = P(\mathbf{X}|\mathbf{B})$. We now show that the refined probability function Pr' determined by P' is such that $Pr'(\texttt{A} \rightarrow_{\texttt{m}} \texttt{B}) = 1$.

> Now, the Pr'-non-zero sequences are those in $\Uparrow(\mathbf{BEL} \cap \mathbf{B})$, which rank worlds in $\mathbf{BEL} \cap \mathbf{B}$ above any not in that state:
>
> $$\Uparrow(\mathbf{BEL} \cap \mathbf{B}) = \begin{Bmatrix} \langle w_1, w_3, w_2 \rangle \\ \langle w_3, w_1, w_2 \rangle \end{Bmatrix}$$
>
> Notice that the falsemaking $\langle w_3, w_2, w_1 \rangle$ is not among these. Indeed, we know that $\texttt{A} \rightarrow_{\texttt{m}} \texttt{B}$ is true at both sequences in $\Uparrow(\mathbf{BEL} \cap \mathbf{B})$:
>
> True at $\langle w_1, w_3, w_2 \rangle$ since its first $m^{\mathbf{A}}_{w_1}$-worlds is $w_1 \in \mathbf{B}$.
> True at $\langle w_3, w_1, w_2 \rangle$ since its first $m^{\mathbf{A}}_{w_3}$-world is $w_1 \in \mathbf{B}$.

Again, it bears emphasizing exactly *why* we predict this result. Conditionalizing a refined probability measure *Pr* on some *Pr*-non-zero refined content will sometimes generate a measure that encodes the probabilistic inferential dispositions of no possible agent! But, conditionalizing a probability function P on any P-non-zero proposition will generate a probability measure of some agent—the measure that agent is disposed to adopt given that proposition. And for this latter function P', its refined counterpart Pr' will model the probabilistic inferential dispositions of the agent whose probability function is P'.

This result is another upshot of the violations of Conditionalization^{-} predicted in Chapter 3:

> **Conditionalization^{-}**
>
> To come to believe A is at most to rule out all possibilities in which A is false.

We have just seen why this is. In coming to believe **B**, we also come to believe $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ (where m is bounded by our prior belief state and we previously regarded **A** as possible), since learning **B** changes our inferential dispositions just enough to believe the latter. Thus, the non-factualist contextualist theory from Chapter 3 predicts exactly the violations of Conditionalization^{-} that are needed to falsify Closure Under Conditionalization and thus avoid the trivialization of Agential Stalnaker's Thesis by Lewis's triviality proof.[5]

Summary: *Diachronic Triviality*

Lewis's triviality results provide further independent grounds for giving up Conditionalization^{-}; here, our theory avoids the results by jettisoning Closure Under Conditionalization—not all refined probability functions in the class closed under conditionalization model the probabilistic inferential dispositions of any agent. Doing so allows us to preserve Agential Stalnaker's Thesis in light of the diachronic triviality results.

[5] Exactly the same move will work to avoid the generalizations of Lewis due to Bradley 2000, 2007; Milne 2003; Hájek 2011b. See Khoo & Santorio 2018 for discussion.

5.3 Synchronic Triviality

In the next two sections, I will consider the case for further restricting Agential Stalnaker's Thesis to rule out instances involving right- and left-nested conditionals:

(7) a. $A \to (B \to C)$ — Right-nested conditional
b. $(A \to B) \to C$ — Left-nested conditional

5.3.1 Right-Nested Conditionals

Recall the problem with Lewis's proof arose at the step in which it established:

(L1) $Pr(\mathtt{A} \to_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = 1$

(L2) $Pr(\mathtt{A} \to_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = 1$

The proof assumed Closure Under Conditionalization to get to this result, which we have found independent reason to reject. However, Fitelson 2015 shows how to derive (L1)/(L2) without such an assumption; indeed, Fitelson shows how to derive (L1)/(L2) using only instances of Stalnaker's Thesis for a single probability function. Hence, Fitelson's triviality result is **synchronic**. Fitelson's derivation appeals to the validity of Import Export:

Import Export

$\vDash A \to (B \to C) \equiv (A \wedge B) \to C$

In particular, that its instances are probabilistically equivalent:

Probabilistic Import Export

$Pr(\mathtt{A} \to_{\mathtt{m}} (\mathtt{B} \to_{\mathtt{m}} \mathtt{C})) = Pr((\mathtt{A} \wedge \mathtt{B}) \to_{\mathtt{m}} \mathtt{C})$

This principle seems quite plausible—sentences of these forms do sound equivalent:

(8) a. If a Republican won the election, then if Reagan did not win, Anderson did.
b. If a Republican won the election and Reagan did not win, Anderson did.

The proof of (L1) from Probabilistic Import Export is straightforward. We assume only two instances of Agential Stalnaker's Thesis. Suppose $Pr(\mathtt{AB}) > 0$ and assume m is bounded by Pr throughout.

i. $Pr(\mathtt{B} \rightarrow_{\mathtt{m}} (\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B})) = Pr((\mathtt{B} \wedge \mathtt{A}) \rightarrow_{m} B)$ Probabilistic Import Export
ii. $Pr((\mathtt{B} \wedge \mathtt{A}) \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{AB}) = 1$ Agential Stalnaker's Thesis
iii. $Pr(\mathtt{B} \rightarrow_{\mathtt{m}} (\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B})) = 1$ (i), (ii)
iv. $Pr(\mathtt{B} \rightarrow_{\mathtt{m}} (\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B})) = Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B})$ Agential Stalnaker's Thesis
v. $Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}|\mathtt{B}) = 1$ (iii), (iv)

However, here appearances are misleading. In fact, my semantics as stated does not predict Import Export, nor does it predict Probabilistic Import Export.[6] To predict Import Export (and thus Probabilistic Import Export), I decompose the semantic contribution of conditionals into two parts: one, a shifty conditional operator $\rightarrow$, that merely shifts the modal base of embedded clauses, and two, a covert selection function modal $\triangleright$, that contributes the modal force of conditionals (cf. Lewis 1975; Kratzer 1981, 1986, 1991; Gillies 2009, 2010; Cariani & Santorio 2018; Khoo & Mandelkern 2019):

Shifty Conditionals
$[\![A \rightarrow B]\!]^{m,\sigma} = 1$ iff $[\![B]\!]^{m^{\mathbf{A}},\sigma} = 1$.
(Where $m^{\mathbf{A}}_{\sigma} = m_{\sigma} \cap \mathbf{A}$, for all σ)

Covert Modal
$[\![\triangleright A]\!]^{m,\sigma} = 1$ iff $\sigma(m_{\sigma}) \in \mathbf{A}$

[6] Here is a counterexample to Import Export: m is Closed and $m_{w_1} = \{w_1, w_2, w_3\}$, where $w_1 \in \mathbf{ABC}$ and $w_2 \in \mathbf{A\bar{B}C}$ and $w_3 \in \mathbf{\bar{A}B\bar{C}}$. Then, for any $w \in m_w$: for any σ: $\sigma(m^{\mathbf{AB}}_{\sigma}) \in \mathbf{C}$. Next, notice that for any σ such that $\sigma_w = w_2 : \sigma(m^{\mathbf{A}}_{\sigma}) = w_2$ and for some $\sigma' : \sigma'_w = w_2 : \sigma'(m^{\mathbf{B}}_{\sigma'}) = w_3$. But $w_3 \notin \mathbf{C}$. So, we predict that not all σ are such that $[\![(A \rightarrow B) \rightarrow C]\!]^{m,\sigma} = 1$.

I propose that the covert modal appears just over the consequent, and with embedded conditionals there is only one covert modal (cf. Khoo 2013):

(9) a. $A \to \rhd B$
b. $A \to (B \to \rhd C)$

For simple conditionals, we get the same truth conditions as before, but for right-nested conditionals, we derive Import Export (and thus Probabilistic Import Export, which follows):

$$\begin{aligned}&[\![A \to (B \to \rhd C)]\!]^{m,\sigma} = 1 \text{ iff}\\ &\quad [\![B \to \rhd C]\!]^{m^{\mathbf{A}},\sigma} = 1 \text{ iff}\\ &\quad \sigma(m^{\mathbf{AB}}_{\sigma}) \in \mathbf{C} \text{ iff}\\ &\quad [\![\rhd C]\!]^{m^{\mathbf{AB}},\sigma} = 1 \text{ iff}\\ &\quad [\![(A \wedge B) \to \rhd C]\!]^{m,\sigma} = 1\end{aligned}$$

With this updated semantics in hand, we can now diagnose which step of Fitelson's proof we block. To keep track of the shiftiness on the modal induced by the conditional, I will rewrite '$\mathtt{A} \to \rhd_{\mathtt{m}}\mathtt{B}$' as '$\rhd_{m^{\mathbf{A}}}\mathtt{B}$'. Note this is without harm, since:

$$\begin{aligned}&\mathtt{A} \to \rhd_{\mathtt{m}}\mathtt{B} =\\ &\quad \{\sigma\colon [\![A \to \rhd B]\!]^{m,\sigma} = 1\} =\\ &\quad \{\sigma\colon [\![\rhd B]\!]^{m^{\mathbf{A}},\sigma} = 1\} =\\ &\quad \rhd_{m^{\mathbf{A}}}\mathtt{B}\end{aligned}$$

So, take the following instance of Agential Stalnaker's Thesis:

$$Pr(\mathtt{A} \to (\mathtt{B} \to \rhd_{\mathtt{m}}\mathtt{C})) = Pr(\mathtt{B} \to \rhd_{\mathtt{m}}\mathtt{C}|\mathtt{A})$$

Note, given our assumptions, this is equivalent to:

$$Pr(\rhd_{\mathtt{m}^{\mathbf{AB}}}\mathtt{C}) = Pr(\rhd_{\mathtt{m}^{\mathbf{B}}}\mathtt{C}|\mathtt{A})$$

However, we predict these contents are distinct, and hence we predict failures of Agential Stalnaker's Thesis of this kind. Here is an example: $A \to (B \to A)$. The probability of the following content:

$\triangleright_{\mathtt{m}^{\mathbf{AB}}} \mathtt{B}$

(Which, recall, is $\{\sigma : \sigma(m_\sigma^{\mathbf{AB}}) \in \mathbf{B}\}$)

is 1, since the probability of picking an **AB**-world that is a **B**-world is 1. But this comes apart from the probability of:

$\triangleright_{\mathtt{m}^{\mathbf{A}}} \mathtt{B}$

(Which, recall, is $\{\sigma : \sigma(m_\sigma^{\mathbf{A}}) \in \mathbf{B}\}$)

given

$\mathtt{A}$

The probability of the latter equals the probability of picking a **B**-world from the **A**-worlds given that you are at an **A**-world (which, given Strong Centering, just is $P(\mathbf{B}|\mathbf{A})$). But, is this prediction plausible? That is, are there intuitive cases in which the probability of a conditional with a conditional consequent comes apart from the probability of the conditional in its consequent given its antecedent? I think there is actually independent evidence for this. I borrow the following case from Paolo Santorio (see Khoo & Santorio 2018):

(10) If the die landed on an odd, then, if it landed on a number greater than 3, it landed on 5.

O = odd
G = greater than 3
F = five

Supposing the die has six sides and is fair, (10) seems like it should have probability 1. Next, consider the embedded conditional:

(11) If it landed on a number greater than 3, it landed on 5.

$\triangleright_{m^{G}} F$

Intuitively, this conditional has a probability of 1/3. Now, we can calculate the probability of (11) given that the die landed even as follows (note: we assume that the probability of odd is 1/2):

$$Pr(\triangleright_{m^{G}}\mathrm{F}|\mathrm{O}) = \frac{Pr(\mathrm{O}|\triangleright_{m^{G}}\mathrm{F})\cdot\overbrace{Pr(\triangleright_{m^{G}}\mathrm{F})}^{1/3}}{\underbrace{Pr(\mathrm{O})}_{1/2}}$$

Given this equation (which follows from Bayes' Theorem), we know that the upper bound on $Pr(\triangleright_{m^{G}}\mathrm{F}|\mathrm{O})$ is $1/3/1/2 = 2/3$. Thus, we have a case where the probability of $\triangleright_{m^{AB}}\mathrm{C}$ comes apart from the probability of $\triangleright_{m^{A}}\mathrm{B}$ given A.

The reason this all works within our revised semantics is that the content of the conditional $B \rightarrow C$ as it appears embedded under '$A \rightarrow$' is different than when it appears unembedded, owing to the information sensitivity of indicative conditionals that is shifted when embedded under another indicative (cf. Kratzer 1986; Khoo 2013). Thus, we should not expect the probability of sentences of the form $A \rightarrow (B \rightarrow C)$ to equal the conditional probability of $B \rightarrow C$ given A, except in the rare cases in which the latter is equal to the probability of $(A \wedge B) \rightarrow C$ (see the discussion in Khoo & Mandelkern 2019).

The foregoing reasoning also extends to any expression that is shiftable by *if*. This includes modals like *might* and *must*, and adverbial quantifiers like *always* and *usually* (cf. Lewis 1975; von Fintel & Iatridou 2002; Khoo 2011). Thus, the solution to Fitelson's triviality result that I advocate is to reject Stalnaker's Thesis and instead adopt a version restricted to non-shiftable consequents, which includes conditionals. That is the bolded part in the statement of Restricted Agential Stalnaker's Thesis:

Restricted Agential Stalnaker's Thesis

For every *Pr* that encodes the probabilistic inferential dispositions of a possible agent, modal base *m* such that *m* is bounded by *Pr*, and **for all non-conditional sentences** *A* and ***B***: if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

To keep things simple in what follows, I will revert back to the less complex non-shifty semantics in my presentation of the material. Thus, do not be alarmed when you no longer see covert modals represented in the logical forms. This is done on purpose to ease readability.

5.3.2 Left-Nested Conditionals

We have motivated restricting Stalnaker's Thesis to exclude right-nested conditionals. Here, I motivate excluding left-nested conditionals, of the form $(A \rightarrow B) \rightarrow C$. There is a tension between validating Stalnaker's Thesis for left-nested conditionals and the logical principle CSO:[7]

CSO
$A \rightarrow B, B \rightarrow A, A \rightarrow C \vDash B \rightarrow C$

Since CSO is equivalent to Limited Transitivity and Antecedent Disjunction Introduction, this tension extends to those principles as well.

Limited Transitivity (LT)
$A \rightarrow B, (A \wedge B) \rightarrow C \vDash A \rightarrow C$

Antecedent Disjunction Introduction (ADI)
$A \rightarrow C, B \rightarrow C \vDash (A \vee B) \rightarrow C$

van Fraassen 1976; Bacon 2015 validate Stalnaker's Thesis for left-nested conditionals, and Bacon argues that this is worth the cost of invalidating CSO, since the counterexamples all involve left-nested

[7] CSO is valid on our sequence semantics because each sequence is a total order on **W** and thus:

If $\sigma(\mathbf{A}) \in \mathbf{B}$ and $\sigma(\mathbf{B}) \in \mathbf{A}$, then $\sigma(\mathbf{A}) = \sigma(\mathbf{B})$.

conditionals, and we might think that intuitions regarding those conditionals are unusual anyway.[8] However, in this section, I will argue instead that we should violate Stalnaker's Thesis for left-nested conditionals and validate CSO.

Stalnaker 1976 generates a triviality result that relies on CSO, one instance of Stalnaker's Thesis, and assumes a background logic of Strong Centering and Conditional Non-Contradiction:

Strong Centering (SC)
$A \vDash (B \equiv A \to B)$

Conditional Non-Contradiction (CNC)
$\vDash \neg[(A \to B) \wedge (A \to \neg B)]$

To get a sense of how the proof works, here is the reasoning simplified (this follows the presentation in Edgington 1995). Start with a definition:

> *Pr* is non-trivial iff there are sentences A, B such that $Pr(\mathtt{A}) > 0, Pr(\bar{\mathtt{A}}) > 0$ and $0 < Pr(\mathtt{B}|\mathtt{A}) < 1$.

Stalnaker shows that given some non-trivial *Pr* and A, B such that $0 < Pr(\mathtt{A} \to \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A}) < 1$, we can construct a counterexample to (ST). The counterexample is $C \to D$, where:

- $C = A \vee \neg(A \to B)$
- $D = A \wedge \neg B$

Make the simplifying assumption for now that each world is equally likely and that there are only four worlds. Then we can represent the distribution of the truth values of the relevant sentences via the following table:

[8] And thus alleged counterexamples to CSO such as those due to Tichý 1976; Tooley 2002 are not relevant to this issue. See Stalnaker 1984 for discussion.

w	$Pr(\{w\})$	A	B	$A \wedge \neg B$ (=D)	$A \to B$	$A \vee \neg (A \to B)$ (=C)	$C \to D$
1	1/4	T	T	F	T	T	F
2	1/4	T	F	T	F	T	T
3	1/4	F		F	T	F	F
4	1/4	F		F	F	T	F

By Strong Centering, we know that $A \to B$ must be true at w_1 and false at w_2. We stipulate that $A \to B$ is true at w_3 and false at w_4 to ensure that $Pr(\mathtt{A} \to \mathtt{B}) = Pr(B|A) = 1/2$. Similarly, $C \to D$ is false at w_1, w_4 and true at w_2 by Strong Centering. But why is $C \to D$ false at w_3? We reason as follows:

1. $A \to B$	From truth table
2. $(A \vee \neg(A \to B)) \to (A \wedge \neg B)$	Reductio assumption
3. $(A \vee \neg(A \to B)) \to A$	2, $\wedge$-elim
4. $(A \vee \neg(A \to B)) \to \neg B$	2, $\wedge$-elim
5. $A \to A$	Trivial
6. $A \to (A \vee \neg(A \to B))$	5, $\vee$-intro
7. $(A \vee \neg(A \to B)) \to B$	1, 3, 6, CSO
8. $\bot$	4, 7, CNC

Stalnaker's counterexample thus reveals the tension between validating instances of Stalnaker's Thesis for left-nested conditionals and CSO. Given our background logic, we have to give up one, thus either further constraining Stalnaker's Thesis or restricting the scope of CSO.

In favor of validating CSO is the fact that Limited Transitivity seems valid, even for left-nested conditionals. Here are two cases, that both seem valid to me:

(12) a. If the coin landed heads if it was flipped, then it was flipped.
$(A \to B) \to A$

b. If the coin landed heads if it was flipped and was flipped, it was flipped by John.
$(A \to B \wedge A) \to C$

c. So, if the coin landed heads if it was flipped, it was flipped by John.
$(A \rightarrow B) \rightarrow C$

(13) a. If the dog barked if he was hungry, then someone fed him.
b. If the dog barked if he was hungry and someone fed him, then there is no food left.
c. So, if the dog barked if he was hungry, then there is no food left.

Similarly, instances of Antecedent Disjunction Introduction involving left-nested conditionals seem valid as well:

(14) a. If vase #1 broke if it was dropped, one of the vases was fragile.
b. If vase #2 broke if it was dropped, one of the vases was fragile.
c. So, if either vase #1 broke if it was dropped or vase #2 broke if it was dropped, one of the vases was fragile.

(15) a. If Smith bought two tickets if she bought any, she and Jones are both going to the show.
b. If Smith called Jones if her phone was charged, she and Jones are both going to the show.
c. So, if either Smith bought two tickets if she bought any or she called Jones if her phone was charged, she and Jones are both going to the show.

The fact that these inferences seem valid suggests that CSO is valid in full generality. But then we expect there to be counterexamples to Stalnaker's Thesis involving left-nested conditionals. Can we find such cases? Start with the following scenario:

Poker Paul. Paul is playing poker against Nancy. Nancy has a weak hand, but it is still possible Paul's is weaker. It is also possible for Paul to win even if he has the weaker hand, but since Nancy is a good player, that is unlikely. Cheating is a way to increase your success of winning with a weaker hand, and Paul is not opposed to cheating, though it is unlikely he cheated if he had the better hand. But, if Paul won with the weaker hand, it is overwhelmingly likely that he cheated. Now consider:

(16) If Paul won if he had the weaker hand, he cheated.

Given the information above, how likely is (16)? Would you say it is unlikely, fifty-fifty, likely, or perhaps you don't have enough information? To me, there is a very clear intuition that (16) is very likely. But now consider a slight variation on the case to streamline judgments. Suppose all of the above is true, except for the following: it is actually quite likely that Paul will win even if he has the weaker hand, since Nancy is a weak player (maybe Paul is really good at bluffing). Do your intuitions change? My intuition is that (16) gets less likely as we get more evidence that Paul won even if he had the weaker hand. In the first case, Paul's winning if he had the weaker hand most likely had to do with his cheating (since he was likely not good enough to beat Nancy with a weaker hand); but in the second case, Paul's winning if he had the weaker hand was less likely a result of his cheating and more likely a case of being good enough to beat Nancy with a weaker hand.

However, the corresponding conditional probability that Paul cheated, assuming he won if he had the weaker hand, is low. To see this, we need to put some numbers on the example. Let H = *Paul had the weaker hand*, W = *Paul won*, and C = *Paul cheated*:

- $Pr(\mathtt{H}) = 0.25$ (Paul likely does not have the weaker hand)
- $Pr(\mathtt{W}|\mathtt{H}) = 0.2$ (Paul likely did not win if he had the weaker hand)
- $Pr(\mathtt{C}|\bar{\mathtt{H}}) = 0.2$ (Paul likely did not cheat if he did not have the weaker hand)
- $Pr(\mathtt{C}|\mathtt{HW}) = 0.95$ (Paul likely cheated if he had the weaker hand and won)
- $Pr(\mathtt{C}|\mathtt{H} \rightarrow \mathtt{W}) = \underbrace{Pr(\mathtt{C}|\mathtt{HW})}_{0.95} \cdot \underbrace{Pr(\mathtt{H})}_{0.25} + \underbrace{Pr(\bar{\mathtt{H}}\mathtt{C})}_{0.15} = 0.3875$

- Calculation in footnote.[9]
- $Pr(\bar{\mathrm{H}}\mathrm{C}) = \underbrace{Pr(\bar{\mathrm{H}})}_{0.75} \cdot \underbrace{Pr(\mathrm{C}|\bar{\mathrm{H}})}_{0.2} = 0.15$

This value is intuitively too low.

Another counterexample in a slightly different vein was suggested by Edgington 1991:

> **Edgington's Coin**. We have a coin, x, which is either double-headed or double-tailed. We are fifty-fifty as to each possibility. Similarly, we think it is 0.5 likely that x was flipped, and thus think it is 0.5 likely that x landed heads given that it was flipped.

Given this, how likely is it that:

(17) If the coin landed heads if it was flipped, it was double-headed.

Intuition suggests we should be certain of this, because the probability that the coin was double-headed given that it was flipped and landed heads is 1. That's Edgington's intuition, too. But she points out that then the probability of (17) can't equal the corresponding conditional probability, which is 0.75:

[9]
$$
\begin{aligned}
Pr(\mathrm{C}|\mathrm{H} \to \mathrm{W}) &= \frac{Pr(\mathrm{H} \to \mathrm{W}|\mathrm{C}) \cdot Pr(\mathrm{C})}{Pr(\mathrm{H} \to \mathrm{W})} \\
&= \frac{[Pr(\mathrm{H} \to \mathrm{W}|\mathrm{CH}) \cdot Pr(\mathrm{H}|\mathrm{C}) + Pr(\mathrm{H} \to \mathrm{W}|\mathrm{C}\bar{\mathrm{H}}) \cdot Pr(\bar{\mathrm{H}}|\mathrm{C})] \cdot Pr(\mathrm{C})}{Pr(\mathrm{H} \to \mathrm{W})} \\
&= \frac{[Pr(\mathrm{W}|\mathrm{CH}) \cdot Pr(\mathrm{H}|\mathrm{C}) + Pr(\mathrm{W}|\mathrm{H}) \cdot Pr(\bar{\mathrm{H}}|\mathrm{C})] \cdot Pr(\mathrm{C})}{Pr(\mathrm{W}|\mathrm{H})} \\
&= \frac{Pr(\mathrm{W}|\mathrm{CH}) \cdot Pr(\mathrm{H}|\mathrm{C}) \cdot Pr(\mathrm{C}) + Pr(\mathrm{W}|\mathrm{H}) \cdot Pr(\bar{\mathrm{H}}|\mathrm{C}) \cdot Pr(\mathrm{C})}{Pr(\mathrm{W}|\mathrm{H})} \\
&= \frac{Pr(\mathrm{HCW}) + Pr(\mathrm{W}|\mathrm{H}) \cdot Pr(\bar{\mathrm{H}}\mathrm{C})}{Pr(\mathrm{W}|\mathrm{H})} \\
&= \frac{Pr(\mathrm{W}|\mathrm{H}) \cdot Pr(\mathrm{H}) \cdot Pr(\mathrm{C}|\mathrm{HW}) + Pr(\mathrm{W}|\mathrm{H}) \cdot Pr(\bar{\mathrm{H}}\mathrm{C})}{Pr(\mathrm{W}|\mathrm{H})} \\
&= Pr(\mathrm{C}|\mathrm{HW}) \cdot Pr(\mathrm{H}) + Pr(\bar{\mathrm{H}}\mathrm{C})
\end{aligned}
$$

$$Pr(\mathtt{D_H}|\mathtt{F} \rightarrow \mathtt{H}) = \underbrace{Pr(\mathtt{D_H}|\mathtt{FH})}_{1} \cdot \underbrace{Pr(\mathtt{F})}_{0.5} + \underbrace{Pr(\bar{\mathtt{F}}\mathtt{D_H})}_{0.25} = 0.75$$

Examples like these motivate further restricting Stalnaker's Thesis to exclude left-nested conditionals:

Restricted Agential Stalnaker's Thesis

For every *Pr* that encodes the probabilistic inferential dispositions of a possible agent, modal base *m* such that *m* is bounded by *Pr*, and **for all non-conditional** *A* and *B*: if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

But what then should we aim to predict about the probabilities of left-nested conditionals? Edgington suggests that $Pr((\mathtt{A} \rightarrow \mathtt{B}) \rightarrow \mathtt{C}) = Pr(\mathtt{C}|\mathtt{BA})$. Indeed, this predicts the intuitively correct probability for (17), as well as the intuitive high probability for (16). However, assuming Strong Centering, this view wrongly predicts that $Pr((\mathtt{A} \rightarrow \mathtt{B}) \rightarrow (\mathtt{B} \rightarrow \mathtt{A})) = 1$. That this wrong can be seen from examples like:

(18) If the vase broke if dropped, then it was dropped if it broke.

This conditional may be false, owing to the fact that the vase was fragile (and hence broke if dropped) but could have broken in some way other than being dropped (perhaps by being hit with a hammer).

Turning back to the theory motivated so far, as it stands, we predict that left-nested conditionals are undefined, since our Baseline Semantics requires that conditional antecedents are propositional. However, we can generate a propositional content out of a conditional to embed in the sequence function in a way that invalidates Stalnaker's Thesis for left-nested conditionals, but also doesn't make the wrong prediction regarding (18). The strategy is to define the propositional content pointwise by each σ as follows. Let $[A \rightarrow B]^{m,\sigma} = \mathbf{AB} \cup \{w \in \bar{\mathbf{A}} : \sigma(m_\sigma \cap \mathbf{A}) \in \mathbf{B}\}$. This is either the set of $\mathbf{A} \supset \mathbf{B}$-worlds (if $\sigma(m_\sigma \cap \mathbf{A})$

$\in$ **B**) or the set of **AB**-worlds (otherwise). Then, we let these propositions pointwise determine the content of $(A \to B) \to C$ as follows:[10]

Pointwise
$[\![(A \to B) \to C]\!]^{m,\sigma} = 1$ iff $\sigma(m_\sigma \cap [A \to B]^{m,\sigma}) \in \mathbf{C}$.

This view invalidates Stalnaker's Thesis for left-nested conditionals.

Here is a simple model. Suppose $Pr(\mathtt{B}|\mathtt{A}) = 1$. Then, **Pointwise** predicts that $Pr((\mathtt{A} \to \mathtt{B}) \to \mathtt{C}) = Pr(\mathtt{C}|\mathtt{A} \supset \mathtt{B}) = Pr(\mathtt{C})$. But we still have:

$$Pr(\mathtt{C}|\mathtt{A} \to \mathtt{B}) = Pr(\mathtt{C}|\mathtt{AB}) \cdot Pr(\mathtt{A}) + Pr(\bar{\mathtt{A}}\mathtt{C})$$

And this latter value can clearly come apart from $Pr(\mathtt{C})$ even if $Pr(\mathtt{B}|\mathtt{A}) = 1$. For instance, even supposing $Pr(\mathtt{C}|\mathtt{AB}) = 1$, it predicts:

$$Pr(\mathtt{C}|\mathtt{A} \to \mathtt{B}) = Pr(\mathtt{A}) + Pr(\bar{\mathtt{A}}\mathtt{C})$$

and then as long as $Pr(\mathtt{AC}) \neq Pr(\mathtt{A})$, the values diverge.

In Appendix B to this chapter, I fully calculate the probabilities for left-nested conditionals predicted by **Pointwise**. For now, I think we have shown that the strategy is at least a viable one to pursue.

Summary: *Synchronic triviality*

The synchronic triviality results of Stalnaker and Fitelson motivate further restricting Agential Stalnaker's Thesis to only conditionals that do not contain conditionals embedded either in their antecedents or consequents.

[10] I am grateful to Arc Kocurek for suggesting this strategy to me.

5.4 Tenability

Our goal now is to extend probability measures over propositions to measures over refined contents, allowing us to calculate the probabilities of the latter in the usual way. This final section is the most technical of the book, and readers unfamiliar or uninterested in the formal proof may skip this part without loss of comprehension for what follows. The key thing to note is that Restricted Agential Stalnaker's Thesis is provably tenable, meaning that there are non-trivial models on which it holds in full generality:

Restricted Agential Stalnaker's Thesis

For every Pr that encodes the probabilistic inferential dispositions of a possible agent, modal base m such that m is bounded by Pr, and **for all non-conditional** A and B: if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

As mentioned above, my inspiration here is van Fraassen 1976. Start with a mass function p that assigns probabilities to worlds, which is normalized on **BEL**:

p is normalized on **BEL** iff for all $w \in \mathbf{BEL} : p(w) > 0$ and for all $w' \notin \mathbf{BEL} : p(w') = 0$.

We can define P in terms of p as follows:

Probability

$$P(\mathbf{A}) = \sum_{w \in \mathbf{A}} p(w)$$

As a shorthand for referring to certain sets of sequences, let $\Sigma[w_i^1, \ldots, w_j^n]$ be the set of sequences whose first world is $w_i, \ldots,$ and whose nth world is w_j. Thus, for instance, we have:

$$\Sigma[w_a^1, w_b^2] = \begin{Bmatrix} \langle w_a, w_b, w_c, w_d, \dots \rangle \\ \langle w_a, w_b, w_d, w_c, \dots \rangle \\ \langle w_a, w_b, w_e, w_c, \dots \rangle \\ \vdots \end{Bmatrix}$$

We can now define a probability measure over sets of sequences Q in terms of p as follows:

(Q1) For any $w \in \mathbf{BEL}$:
$$Q(\Sigma[w^1]) = p(w^1)$$

This says that the probability of a sequence set starting with w^1 just is the probability that w^1 is actual.

(Q2) For any $w^1, w^2 \in \mathbf{BEL}$:
$$Q(\Sigma[w^1, w^2]) = \frac{p(w^2)}{P(\mathbf{BEL} \setminus \{w^1\})} \cdot p(w^1)$$

This says that the probability of a sequence set starting $\langle w^1, w^2, \dots \rangle$ just is the probability of randomly picking w^1 and then w^2 (in that order) from **BEL**.

(Q3) For all $w^1, \dots, w^n \in \mathbf{BEL}$:
$$Q(\Sigma[w^1, \dots, w^n]) = \frac{p(w^n)}{P(\mathbf{BEL} \setminus \{w^1, \dots, w^{n-1}\})} \cdot Q(\Sigma[w^1, \dots, w^{n-1}])$$

This defines Q for all sequence sets whose sequences rank worlds in **BEL** before worlds not in **BEL**. For now, we will not extend the definition to assign probabilities to unique sequences, since we do not have to in order to generate the tenability result for indicative conditionals (however, we will for subjunctive conditionals—a topic I come back to in Chapter 10).

Let us look at an example to see this definition in action.

- $\mathbf{W} = \{w_1, w_2, w_3, w_4, w_5\}$
- $\mathbf{BEL} = \{w_1, w_2, w_3\}$
- $p(w_1) = p(w_2) = p(w_3) = 1/3$

- $\Uparrow\mathbf{BEL} = \begin{Bmatrix} \sigma_1 : \langle w_1, w_2, w_3, w_4, w_5\rangle & \sigma_5 : \langle w_2, w_1, w_3, w_4, w_5\rangle & \sigma_9 : \langle w_3, w_1, w_2, w_4, w_5\rangle \\ \sigma_2 : \langle w_1, w_2, w_3, w_5, w_4\rangle & \sigma_6 : \langle w_2, w_1, w_3, w_5, w_4\rangle & \sigma_{10} : \langle w_3, w_1, w_2, w_5, w_4\rangle \\ \sigma_3 : \langle w_1, w_3, w_2, w_4, w_5\rangle & \sigma_7 : \langle w_2, w_3, w_1, w_4, w_5\rangle & \sigma_{11} : \langle w_3, w_2, w_1, w_4, w_5\rangle \\ \sigma_4 : \langle w_1, w_3, w_2, w_5, w_4\rangle & \sigma_8 : \langle w_2, w_3, w_1, w_5, w_4\rangle & \sigma_{12} : \langle w_3, w_2, w_1, w_5, w_4\rangle \end{Bmatrix}$

Here are some probability assignments. You can verify that these correspond to the relevant proportion of $\sigma \in \Uparrow\mathbf{BEL}$.

$$Q(\Sigma[w_1]) = p(w_1) = 1/3$$

$$Q(\Sigma[w_1, w_2]) = \frac{p(w_2)}{P(\mathbf{BEL} - \{w_1\})} \cdot p(w) = {}^{1/3}/_{2/3} \cdot 1/3 = 1/6$$

$$Q(\Sigma[w_1, w_2, w_3]) = \frac{p(w_3)}{P(\mathbf{BEL} - \{w_1, w_2\})} \cdot Q(\Sigma[w_1, w_2]) = 1 \cdot 1/6 = 1/6$$

This will hold for all Σ containing sequences that start with w_1, w_2, w_3 in any order (since the worlds in **BEL** are equiprobable). Thus, for instance, we have:

$$\Sigma[w_2, w_1, w_3] = \{\sigma_5, \sigma_6\}$$

$$Q(\Sigma[w_2, w_1, w_3]) = 1/6$$

Notice now, for this example at least, we predict that $Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = P(\mathbf{B}|\mathbf{A})$, as long as m is bounded by **BEL**. Suppose that $\mathbf{A} = \{w_1, w_2\}$ and $\mathbf{B} = \{w_1\}$ and that $m_{w_1} = m_{w_2} = m_{w_3} = \{w_1, w_2, w_3\}$. Then we have:

- $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is true at $\sigma_1, \ldots, \sigma_4$ by Strong Centering.
- $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is false at $\sigma_5, \ldots, \sigma_8$ by Strong Centering.
- $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is true at σ_9, σ_{10} because at both $\sigma(m_\sigma \cap \mathbf{A}) = w_1$ and $w_1 \in \mathbf{B}$.
- Thus, $\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}$ is true at exactly one-half of the sequences in $\Uparrow\mathbf{BEL}$ and each has the same probability, so $Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = 1/2 = P(\mathbf{B}|\mathbf{A})$.

Now, it remains to be shown that this holds in general, which is the task I turn to next.

To begin, assume that A, B are non-conditional and that m is bounded by Pr and that $Pr(\mathtt{A}) > 0$. We begin by expanding across $\{\mathtt{A}, \bar{\mathtt{A}}\}$:

$$Pr(\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}) = Pr(\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B} | \mathtt{A}) \cdot Pr(\mathtt{A}) + Pr(\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B} | \bar{\mathtt{A}}) \cdot Pr(\bar{\mathtt{A}})$$

By Strong Centering, we know that

$$Pr(\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B} | \mathtt{A}) = Pr(\mathbf{B} | \mathbf{A})$$

Thus, Restricted Agential Stalnaker's Thesis follows from:

Target

$$Pr(\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B} | \bar{\mathtt{A}}) = P(\mathbf{B} | \mathbf{A})$$

Let $\Sigma[w]_n^{\mathbf{A}}$ be the set of w-led sequences whose first **A**-world in **BEL** is their nth. Then, since m is bounded by Pr and Pr is normalized on **BEL**, Target follows from:

The Key Lemma

$$\forall w \in \mathbf{BEL} \cap \bar{\mathbf{A}} : \forall n : 1 < n \leq |\mathbf{BEL}| - |\mathbf{A}| + 1:$$
$$Q(\Sigma[w]_n^{\mathbf{AB}} | \Sigma[w]_n^{\mathbf{A}}) = P(\mathbf{B} | \mathbf{A})$$

This is because $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ is the set of sequences whose first **A**-world in m is a **B**-world. Since m is bounded by **BEL** and each sequence in $\Uparrow$**BEL** ranks **BEL**-worlds above $\overline{\mathbf{BEL}}$-worlds, the sequences in $\Uparrow$**BEL** at which $\mathtt{A} \rightarrow_{\mathrm{m}} \mathtt{B}$ holds are those whose first **A**-world in **BEL** is a **B**-world. The $\bar{\mathbf{A}}$-led sequences can be partitioned into n groups, according to whether their first **A**-world in **BEL** is their nth. Then, what the Key Lemma states is that, for each $\bar{\mathbf{A}}$-world w in **BEL** and each n, the probability of the set of sequences whose first **AB**-world is their nth (=$\Sigma[w]_n^{\mathbf{AB}}$) given the sequence set whose first **A**-world is their nth

$(=\Sigma[w]^{\mathbf{A}}_{n})$ is equal to $P(\mathbf{B}|\mathbf{A})$. But if this holds for all possible places, we might find the first $\mathbf{A}$-world in a sequence in $\Uparrow\mathbf{BEL}$ (that is, for all $n : 1 < n \leq |\mathbf{BEL}| - |\mathbf{A}| + 1$), then it holds generally, and thus Target follows.

We will work toward the Key Lemma by way of some facts:

Fact 1

$$\forall w \in \mathbf{BEL} \cap \bar{\mathbf{A}}:$$
$$Q(\Sigma[w]^{\mathbf{AB}}_{2}|\Sigma[w]^{\mathbf{A}}_{2}) = P(\mathbf{B}|\mathbf{A})$$

This states that, for any $w \in \mathbf{BEL} \cap \bar{\mathbf{A}}$, the probability of a sequence set whose first $(\mathbf{A} \cap \mathbf{BEL})$-world is its second and also a $\mathbf{B}$-world is equal to $P(\mathbf{B}|\mathbf{A})$.

Proof. Let w be an arbitrary world in $\mathbf{BEL} \cap \bar{\mathbf{A}}$. We know that:

$$\begin{aligned}
Q(\Sigma[w]^{\mathbf{A}}_{2}) &= \sum_{w' \in \mathbf{A}} Q(\Sigma[w, w']) \\
&= \sum_{w' \in \mathbf{A}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})} \cdot p(w) \\
Q(\Sigma[w]^{\mathbf{AB}}_{2}) &= \sum_{w' \in \mathbf{AB}} Q(\Sigma[w, w']) \\
&= \sum_{w' \in \mathbf{AB}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})} \cdot p(w)
\end{aligned}$$

Thus, since Q is a probability function, we have:

$$\begin{aligned}
Q(\Sigma[w]^{\mathbf{AB}}_{2}|\Sigma[w]^{\mathbf{A}}_{2}) &= \frac{Q(\Sigma[w]^{\mathbf{AB}}_{2} \cap \Sigma[w]^{\mathbf{A}}_{2})}{Q(\Sigma[w]^{\mathbf{A}}_{2})} \\
&= \frac{Q(\Sigma[w]^{\mathbf{AB}}_{2})}{Q(\Sigma[w]^{\mathbf{A}}_{2})}
\end{aligned}$$

And thus,

$$\frac{Q(\Sigma[w]_2^{\mathbf{AB}})}{Q(\Sigma[w]_2^{\mathbf{A}})} = \frac{\sum_{w' \in \mathbf{AB}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})} \cdot p(w)}{\sum_{w' \in \mathbf{A}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})} \cdot p(w)}$$
$$= \frac{\sum_{w' \in \mathbf{AB}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})}}{\sum_{w' \in \mathbf{A}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})}}$$
$$= \sum_{w' \in \mathbf{AB}} \frac{p(w')}{P(\mathbf{BEL} \setminus \{w_1\})} \cdot \sum_{w' \in \mathbf{A}} \frac{P(\mathbf{BEL} \setminus \{w_1\})}{p(w')}$$
$$= \frac{\sum_{w' \in \mathbf{AB}} p(w')}{\sum_{w' \in \mathbf{A}} p(w')}$$
$$= \frac{P(\mathbf{AB})}{P(\mathbf{A})} = P(\mathbf{B}|\mathbf{A})$$

Showing that the reasoning is the same for each $n : 1 < n \leq |\mathbf{BEL}| - |\mathbf{A}| + 1$ establishes the Key Lemma.

Proof. Let n some some number such that $1 < n \leq |\mathbf{BEL}| - |\mathbf{A}| + 1$ and w an arbitrary $\mathbf{BEL} \cap \bar{\mathbf{A}}$-world. Start with the following facts again:

$$Q(\Sigma[w]_n^{\mathbf{A}}) = \sum_{w,\ldots,w_{n-1} \in \bar{\mathbf{A}}} \sum_{w_n \in \mathbf{A}} Q(\Sigma[w, \ldots, w_{n-1}, w_n])$$
$$= \sum_{w,\ldots,w_{n-1} \in \bar{\mathbf{A}}} \sum_{w_n \in \mathbf{A}} \frac{p(w_n)}{P(\mathbf{BEL} \setminus \{w, \ldots, w_{n-1}\})}$$
$$\cdot Q(\Sigma[w, \ldots, w_{n-1}])$$

$$Q(\Sigma[w]_n^{\mathbf{AB}}) = \sum_{w,\ldots,w_{n-1} \in \bar{\mathbf{A}} \cup \mathbf{A}\bar{\mathbf{B}}} \sum_{w_n \in \mathbf{AB}} Q(\Sigma[w, \ldots, w_{n-1}, w_n])$$
$$= \sum_{w,\ldots,w_{n-1} \in \bar{\mathbf{A}} \cup \mathbf{A}\bar{\mathbf{B}}} \sum_{w_n \in \mathbf{AB}} \frac{p(w_n)}{P(\mathbf{BEL} \setminus \{w, \ldots, w_{n-1}\})}$$
$$\cdot Q(\Sigma[w, \ldots, w_{n-1}])$$

As above with the proof of Fact 1, we have:

$$Q(\Sigma[w]_n^{\mathbf{AB}}|\Sigma[w]_n^{\mathbf{A}}) = \frac{Q(\Sigma[w]_n^{\mathbf{AB}} \cap \Sigma[w]_n^{\mathbf{A}})}{Q(\Sigma[w]_n^{\mathbf{A}})}$$

Thus,

$$\frac{Q(\Sigma[w]_n^{\mathbf{AB}} \cap \Sigma[w]_n^{\mathbf{A}})}{Q(\Sigma[w]_n^{\mathbf{A}})} =$$

$$\frac{\sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} \frac{p(w_n)}{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})} \cdot Q(\Sigma[w,\ldots,w_{n-1}])}{\sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} \frac{p(w_n)}{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})} \cdot Q(\Sigma[w,\ldots,w_{n-1}])} =$$

$$\frac{\sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} \frac{p(w_n)}{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})}}{\sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} \frac{p(w_n)}{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})}} =$$

$$\sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{AB}} \frac{p(w_n)}{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})} \cdot \sum_{w,\ldots,w_{n-1}\in\bar{\mathbf{A}}}\sum_{w_n\in\mathbf{A}} \frac{P(\mathbf{BEL}\setminus\{w,\ldots,w_{n-1}\})}{p(w_n)} =$$

$$\frac{\sum_{w_n\in\mathbf{AB}} p(w_n)}{\sum_{w_n\in\mathbf{A}} p(w_n)} = P(\mathbf{B}|\mathbf{A})$$

This result holds only for conditionals whose modal base m is bounded by Pr where $Pr(\mathtt{A}) > 0$. But that is a good result, in light of our discussion in §4.1. In Chapter 10, I return to show how to extend this result to predict non-trivial probabilities for subjunctive conditionals.

Appendix A: Trivalent Approaches to Triviality

Here, I consider the possibility of reconciling Stalnaker's Thesis and Non-Triviality by rejecting the Law of Total Probability. Lassiter 2020 argues that we ought to give up this principle as applied to conditionals, and he shows how the following trivalent semantics for conditionals invalidates it, and is therefore able to reconcile Stalnaker's Thesis with Non-Triviality (see also Rothschild 2014; Bradley 2002):

Trivalent Theory

$$A \to B \text{ is } \begin{Bmatrix} \text{true} & \text{if } AB \text{ is true} \\ \text{false} & \text{if } A\bar{B} \text{ is true} \\ \text{undefined} & \text{otherwise} \end{Bmatrix}$$

Lassiter draws on this theory to motivate an alternative definition of sentential probability as the probability that a sentence is true given that it is defined (here, I follow Lassiter in working with a simplified probability function that assigns probabilities to sentences, and ignore relativization to modal bases):

$$P(A) = \frac{P(\mathbf{True}(A))}{P(\mathbf{Def}(A))}, \text{ if } P(\mathbf{Def}(A)) > 0$$

Given this notion of sentential probability, the Triviality Theory immediately predicts Stalnaker's Thesis in full generality:

$$P(A \to B) = \frac{P(\mathbf{True}(A \to B))}{P(\mathbf{Def}(A \to B))} = \frac{P(AB)}{P(A)}, \text{ if } P(A) > 0$$

Furthermore, this definition of sentential truth similarly motivates an alternative definition of conditional probability as follows:

$$P(B|A) = \frac{P(\mathbf{True}(AB))}{P(\mathbf{Def}(B) \wedge \mathbf{True}(A))}$$

And given this definition, we immediately predict (L1) and (L2):

$$P(A \to B|B) = \frac{Pr(\mathbf{True}(A \to B \wedge B))}{P(\mathbf{Def}(A \to B) \wedge \mathbf{True}(B))} = \frac{Pr(AB)}{Pr(AB)} = 1$$

$$P(A \to B|\bar{B}) = \frac{Pr(\mathbf{True}(A \to B \wedge \bar{B}))}{P(\mathbf{Def}(A \to B) \wedge \mathbf{True}(\bar{B}))} = \frac{Pr(AB \wedge \bar{B})}{Pr(A\bar{B})} = 0$$

and thus violations of the Law of Total Probability, since:

$$P(A \rightarrow B) = \underbrace{P(A \rightarrow B|B)}_{1} \cdot P(B) + \underbrace{P(A \rightarrow B|\bar{B})}_{0} \cdot P(\bar{B}) \neq P(B)$$

This is an interesting alternative strategy to explore in more detail. However, trivalent views of this kind notoriously run into trouble accounting for conjunctions and disjunctions of conditionals (see Bradley 2002). More pressing, however, for our purposes, is that this strategy seems to incorrectly predict that $P(A \rightarrow B|B \rightarrow A) = 1$ (its implausibility owing to the non-equivalence of $A \rightarrow B$ and $B \rightarrow A$):

$$\begin{aligned} P(A \rightarrow B|B \rightarrow A) &= \frac{Pr(\mathbf{True}(A \rightarrow B \wedge B \rightarrow A))}{P(\mathbf{Def}(A \rightarrow B) \wedge \mathbf{True}(B \rightarrow A))} \\ &= \frac{Pr(AB)}{Pr(AB)} = 1 \end{aligned}$$

Thus, for now, I will set aside this strategy, although I think it is a worthy one to explore in more detail.

Appendix B: Calculating Left-Nested Probabilities

Recall Pointwise:

Pointwise
$[\![(A \rightarrow B) \rightarrow C]\!]^{m,\sigma} = 1$ iff $\sigma(m_\sigma \cap [A \rightarrow B]^{m,\sigma}) \in \mathbf{C}$

and:

$$[A \rightarrow B]^{m,\sigma} = \begin{cases} \mathbf{A} \supset \mathbf{B} & \text{if } \sigma(m_\sigma \cap \mathbf{A}) \in \mathbf{B} \\ \mathbf{AB} & \text{otherwise} \end{cases}$$

I start by showing that we immediately predict $Pr((\mathtt{A} \rightarrow \mathtt{B}) \rightarrow (\mathtt{B} \rightarrow \mathtt{A}))$ can be less than 1, and thus Pointwise offers an improvement over the predictions of the trivalent theory.

Again suppose $Pr(\texttt{B}|\texttt{A}) = 1$. Then, $Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow (B \rightarrow A)) = Pr((\texttt{A} \supset \texttt{B}) \rightarrow (\texttt{B} \rightarrow \texttt{A}))$, which, by our treatment of right-nested conditionals (see §5.3.1) equals $Pr((\texttt{A} \supset \texttt{B} \wedge \texttt{B}) \rightarrow \texttt{A}))$. But this latter value need not equal 1.

Here is how to calculate $Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C})$ generally, given Pointwise. We expand the probability of the conditional across the partition $\{\texttt{A}, \bar{\texttt{A}}\}$:

$$Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}) =$$
$$\text{(i)}\, Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}|\texttt{A}) \cdot Pr(\texttt{A}) +$$
$$\text{(ii)}\, Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}|\bar{\texttt{A}}) \cdot Pr(\bar{\texttt{A}})$$

We expand (i) as follows. Let σ be an arbitrary **A**-led sequence. Either it is an **AB**-led sequence or an $\mathbf{A\bar{B}}$-led sequence. If the former, then $\texttt{A} \rightarrow \texttt{B}$ is true at σ (by Strong Centering), and $(\texttt{A} \rightarrow \texttt{B}) \rightarrow C$ is thus true iff $\texttt{C}$ is true. If the latter, then $\texttt{A} \rightarrow \texttt{B}$ is false at σ (by Strong Centering), in which case $[A \rightarrow B]^{m,\sigma} = \mathbf{AB}$. But then $Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}|\texttt{A}\bar{\texttt{B}}) = Pr(\texttt{C}|\texttt{AB})$. Thus, either way, (i) equals:

(a) $Pr(\texttt{C}|\texttt{AB}) \cdot Pr(\texttt{AB})$

Next, part (ii). Focus on the $\bar{\mathbf{A}}$-led sequences. We divide them into groups. Let $\texttt{X}$ be those $\bar{\mathbf{A}}$-led sequences such that $\sigma(m_\sigma \cap \mathbf{A}) \in \mathbf{B}$. And let $\texttt{Y}$ be those $\bar{\mathbf{A}}$-led sequences such that $\sigma(m_\sigma \cap \mathbf{A}) \notin \mathbf{B}$. At the $\texttt{X}$-sequences, $[A \rightarrow B]^{m,\sigma} = \mathbf{A} \supset \mathbf{B}$, and at the $\texttt{Y}$-sequences, $[A \rightarrow B]^{m,\sigma} = \mathbf{AB}$. Thus, we have:

$$Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}|\bar{\texttt{A}}\texttt{X}) = Pr(\texttt{C}|\texttt{A} \supset \texttt{B})$$

$$Pr((\texttt{A} \rightarrow \texttt{B}) \rightarrow \texttt{C}|\bar{\texttt{A}}\texttt{Y}) = Pr(\texttt{C}|\texttt{AB})$$

Therefore, by total probability we have:

$$Pr((\mathtt{A} \to \mathtt{B}) \to \mathtt{C}|\bar{\mathtt{A}}) =$$
$$\underbrace{Pr((A \to B) \to C|\bar{\mathtt{A}}\mathtt{X})}_{Pr(\mathtt{C}|\mathtt{A}\supset\mathtt{B})} \cdot \underbrace{Pr(\mathtt{X}|\bar{\mathtt{A}})}_{Pr(\mathtt{B}|\mathtt{A})} + \underbrace{Pr((\mathtt{A} \to B) \to C|\bar{\mathtt{A}}\mathtt{Y})}_{Pr(\mathtt{C}|\mathtt{AB})} \cdot \underbrace{Pr(\mathtt{Y}|\bar{\mathtt{A}})}_{Pr(\bar{\mathtt{B}}|\mathtt{A})}$$

And thus (ii) equals:

(b) $[Pr(\mathtt{C}|\mathtt{A} \supset \mathtt{B}) \cdot Pr(\mathtt{B}|\mathtt{A}) + Pr(\mathtt{C}|\mathtt{AB}) \cdot Pr(\bar{\mathtt{B}}|\mathtt{A})] \cdot Pr(\bar{\mathtt{A}})$

In total, we have:

Left-Nested Probabilities
$Pr((\mathtt{A} \to \mathtt{B}) \to \mathtt{C}) =$
$Pr(\mathtt{C}|\mathtt{AB}) \cdot Pr(\mathtt{AB}) + [Pr(\mathtt{C}|\mathtt{A} \supset \mathtt{B}) \cdot Pr(\mathtt{B}|\mathtt{A}) + Pr(\mathtt{C}|\mathtt{AB}) \cdot Pr(\bar{\mathtt{B}}|\mathtt{A})] \cdot$
$Pr(\bar{\mathtt{A}})$

We can now see what this theory predicts regarding the probability of the Poker Paul conditional. Recall:
Poker Paul

- $Pr(\mathtt{H}) = 0.25$ (Paul likely has the weaker hand)
- $Pr(\mathtt{W}|\mathtt{H}) = 0.2$ (Paul likely did not win if he had the weaker hand)
- $Pr(\mathtt{C}|\bar{\mathtt{H}}) = 0.2$ (Paul likely did not cheat if he did not have the weaker hand)
- $Pr(\mathtt{C}|\mathtt{HW}) = 0.95$ (Paul likely cheated if he had the weaker hand and won)
- $Pr((\mathtt{H} \to \mathtt{W}) \to \mathtt{C})$ is intuitively high, since we think it is likely that Paul could not beat Nancy with a weaker hand (reflected in $Pr(\mathtt{W}|\mathtt{H}) = 0.2$).

$$Pr((\mathtt{H} \to \mathtt{W}) \to \mathtt{C}) =$$
(i) $\underbrace{Pr(\mathtt{C}|\mathtt{HW})}_{0.95} \cdot \underbrace{Pr(\mathtt{H})}_{0.25} +$

(ii) $\Big[(\underbrace{Pr(\mathtt{C}|\mathtt{H} \supset \mathtt{W})}_{x} \cdot \underbrace{Pr(\mathtt{W}|\mathtt{H})}_{0.2} + \underbrace{Pr(\mathtt{C}|\mathtt{HW})}_{0.95} \cdot \underbrace{Pr(\bar{\mathtt{W}}|\mathtt{H})}_{0.8}\Big] \cdot \underbrace{Pr(\bar{\mathtt{H}})}_{0.75}$

$$= 0.95 \cdot 0.25 + [0.2x + 0.95 \cdot 0.8] \cdot 0.75$$
$$= 0.2375 + [0.2x + 0.76] \cdot 0.75$$

We can see that even if $x = 0$, the minimum value we predict is still high ($= 0.8075$).

6

Partition Dependence

We come now to an important complication regarding indicative conditionals. So far, I have defended a theory that predicts the following observations:

Weak Sufficiency*

Coming to accept $A \rightarrow_m B$ (when you do not rule out **AB**) and m is bounded by your beliefs is to come to believe nothing more (no proposition stronger) than that $A \supset B$ is true.

Restricted Agential Stalnaker's Thesis

For every Pr that encodes the probabilistic inferential dispositions of a possible agent, modal base m such that m is bounded by Pr, and for all non-conditional sentences A and B: if $Pr(\mathtt{A}) > 0$,

$$Pr(\mathtt{A} \rightarrow_{\mathtt{m}} \mathtt{B}) = Pr(\mathtt{B}|\mathtt{A})$$

However, it turns out there are counterexamples to both. Recall the Edgington's Coin example from Chapter 5.

Edgington's Coin. Smith has a coin, which is either double-headed or double-tailed (1/2 likely either way). We're not sure whether he flipped the coin, but we know that whether he did is independent of whether it was double-headed or double-tailed.

At this point, there are four kinds of worlds compatible with our information:

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0007

Coin?	Flip?	Landed?
Double-headed	Yes	Heads
Double-headed	No	Neither
Double-tailed	Yes	Tails
Double-tailed	No	Neither

Now suppose that Sue, who has more information, tells us:

(1) If Smith flipped the coin, it landed heads.

What do we thereby learn? It seems we should learn that the coin is double-headed. But notice that this is stronger than the corresponding material conditional $\mathbf{F} \supset \mathbf{H}$. Learning the latter is only enough to rule out line three, and thus not enough to come to learn that the coin is double headed. This is a violation of Weak Sufficiency*. Furthermore, such cases also give rise to counterexamples to Restricted Agential Stalnaker's Thesis as well. Here is the variant:[1]

> **McGee's Coins**. McGee has 10 coins, each given to a different person, A, B, C, ... Five of the coins are double-headed and five are double-tailed. Each person knew the following: if they flip and it lands heads, they win \$1; if they flip and it lands tails, they lose \$.50; nothing happens if they don't flip. Three of the 5 people who received a doubled-headed coin knew their coin was double-headed; while 2 of the 5 people who received a double-tailed coin knew their coin was double-tailed; the rest thought they were flipping a fair coin. Everyone was rational and wanted to win.

Think about person A. Given what we know, we can be certain of the following:

- If A got a double-headed coin and knew it, she flipped and won.
- If A got a double-tailed coin and knew it, she didn't flip.
- If A thought she had a fair coin, she flipped.
  - Since she was rational and flipping maximized her expected utility given her evidence.

[1] Named after Vann McGee, because of McGee 2000, which I discuss below.

Now, how likely is the following:

(2) If A flipped, she won.

Here, there is a natural intuition that this conditional's probability is equal to the probability that A received a double-headed coin, namely 1⁄2. After all, she either received a double-headed coin or a double-tailed coin, and if she flipped a double-headed coin, she won, whereas if she flipped a double-tailed coin, she lost.

If this is right, then we have a counterexample to Restricted Agential Stalnaker's Thesis. The reason is that the conditional probability that A won given that she flipped her coin is 5⁄8. The calculation is straightforward:

- $P(\mathbf{F})$ is equal to the proportion of people who flipped, which is 4⁄5.
    - All five of the double-headers flipped.
    - The two double-tailers who knew their coin didn't flip; the other three who didn't know they had a double-tailed coin did flip.
- $P(\mathbf{FW})$ is equal to the proportion of people who flipped and won, which is 1⁄2.
    - All of the five double-headers flipped and won.
    - No one else flipped and won.
- $P(\mathbf{W}|\mathbf{F}) = \dfrac{P(\mathbf{FW})}{P(\mathbf{F})} = \dfrac{1/2}{4/5} = \dfrac{5}{8}$

I should point out that not everyone shares the judgment that (2) is 1⁄2 likely; some people have the Stalnaker-Thesis intuition that it is 5⁄8 likely. And some people can go back and forth between these judgments. This is a point we'll return to below. For now, I just want to raise the point that there *do* seem to be counterexamples to Restricted Agential Stalnaker's Thesis.[2]

[2] Here is another example (from Steve Yablo, personal communication).

Yablo's Coin. A coin has just been tossed. You know it was weighted 3 to 1 towards tails, so it probably landed on tails. You ask Jones whether the coin landed heads or tails. However, Jones's response is muffled and you cannot

The aim of this chapter is to refine our theory to account for these exceptional cases. Before we do so, however, I want to address a potential objection that the intuitions here are suspect.

In particular, it may be that the Thesis-violating intuitions here are the result of a cognitive illusion akin to base rate neglect—a tendency to ignore general information in favor of specific information (cf. Kahneman & Tversky 1973, 1982; Bar-Hillel 1980). Rothschild 2013 suggests this kind of response. However, if this is correct, then we should expect the mistake to be correctable, perhaps by presenting the information using natural frequencies with respect to an intuitive reference class (cf. Gigerenzer & Hoffrage 1995; Hoffrage et al. 2000). But the Thesis-violating intuitions in McGee's Coin were elicited using natural frequencies rather than probabilities. This casts doubt on the strategy of explaining away these intuitions as base rate neglect.

Sorting out these Thesis-violating intuitions is the topic of this chapter. Building on my work in Khoo 2016, I will argue for a view that predicts these violations as stemming from conditionals being evaluated relative to partitions that are sometimes set by a question under discussion in the context. This view bears several similarities to

> hear her: it sounded more like "heads" than "tails" (although you are not sure). Thus, you conclude that it is likely that:
>
> (i) If Jones is correct, the coin landed heads.
>
> Nonetheless, it seems you should think the following is unlikely:
>
> (ii) If Jones is correct, she said that the coin landed heads.
>
> This is because you still think it likely that the coin landed on tails (since it was weighted towards tails). So, it is more likely that Jones said that it landed tails if she was correct.

Conditional on Jones being correct, "the coin landed heads" is true iff "Jones said that the coin landed heads" is true. So, the probability that the coin landed heads given that Jones is correct is equal to the probability that Jones said "heads" given that Jones is correct. But since (i) is intuitively likely and (ii) intuitively unlikely, at least one of these conditionals (or perhaps both) must be a counterexample to Restricted Agential Stalnaker's Thesis. Notice that, while the intuitions about the probabilities of these conditionals are clear, the corresponding conditional probability judgments are less clear. In particular, if we were to learn Jones is correct, would we think it more likely that she said "heads" and the coin landed heads, or would we think it more likely that she did not say "heads" and the coin landed tails? It is not obvious. For more cases, see McGee 2000; Kaufmann 2004; Moss 2015, 2018; Khoo 2016.

the theory developed independently by Moss 2015, 2018—in particular, Moss also holds that conditionals are evaluated relative to partitions, and her theory is able to predict the Thesis-violating intuitions discussed here. However, as we will see below, our views are actually quite different: where my theory generates the Thesis-violating interpretations by relativizing conditionals to partitions, Moss does so by relativizing both conditionals and probability operators to partitions (and it is the latter relativization that does the work on her theory). As a result, the way we generate Thesis-violations is very different and, interestingly, this leads to different predictions in certain cases. In §6.1, I motivate partition dependence by drawing on insights from Kaufmann 2004. Although I ultimately reject Kaufmann's theory, my theory can be understood as a generalization of his main idea—that conditionals are evaluated relative to partitions. I sketch my proposal for how to integrate partitions into our semantics in §6.2. Although the resulting view predicts a very wide range of possible interpretations for conditionals, in §6.3 I motivate a pragmatic constraint that explains why such Thesis-violating interpretations are generally rare.

6.1 Introducing Partition Dependence

I begin with an observation from Kaufmann 2004. In each of the cases above, there is a background variable $\mathbb{Z}$ that partitions the space of possibilities, cutting across worlds in which the antecedent is true—in the Edgington and McGee coin cases, the partition is whether the coin is double-headed or double-tailed, $\{\mathbf{D_H}, \mathbf{D_T}\}$. If double-headed, the chance that the coin landed heads if flipped is 1, and if double-tailed, the chance that the coin landed heads if flipped is 0.

Kaufmann observes that whenever the chance variable is evidentially (but not causally) dependent on the conditional's antecedent, we will generate violations of Stalnaker's Thesis. Kaufmann conjectures that a conditional $A \rightarrow B$ will violate (Restricted Agential) Stalnaker's Thesis when there is a background chance-determining variable $\mathbb{Z}$ that partitions the space of epistemic possibilities and is such that (see Kaufmann 2004: 595–597):

1. The objective chance of the consequent B depends on its antecedent A and $\mathbb{Z}$,
2. $\mathbb{Z}$ is causally independent of A, and
3. $\mathbb{Z}$ is evidentially dependent on A.

Kaufmann adds to this conjecture a proposal about what the probability of $A \rightarrow B$ will be in such cases. He proposes that conditionals are ambiguous between a local interpretation and a global interpretation. Let '$\rightarrow_{\ell}^{\mathbb{Z}}$' symbolize the local conditional operator (interpreted relative to partition $\mathbb{Z}$), and let '$\rightarrow_{g}$' symbolize the global conditional operator. Kaufmann distinguishes the interpretations by their probabilities as follows:

$$\textbf{Local}: Pr(\texttt{A} \rightarrow_{\ell}^{\mathbb{Z}} \texttt{B}) = \sum_{\mathbf{Z}\in\mathbb{Z}} P(\mathbf{B}|\mathbf{AZ}) \cdot P(\mathbf{Z})$$

$$\textbf{Global}: Pr(\texttt{A} \rightarrow_{g} \texttt{B}) = P(\mathbf{B}|\mathbf{A})$$

Kaufmann predicts that the probability of the local interpretation of $F \rightarrow H$ is equal to the probability that the coin was double-headed in Edgington's Coin:

$$Pr(\texttt{F} \rightarrow_{\ell}^{\mathbb{D}} \texttt{H}) = \underbrace{P(\mathbf{H}|\mathbf{FD_H})}_{1} \cdot \underbrace{P(\mathbf{D_H})}_{1/2} + \underbrace{P(\mathbf{H}|\mathbf{FD_T})}_{0} \cdot \underbrace{P(\mathbf{D_T})}_{1/2} = 1/2$$

Kaufmann's theory promises a major advance on our thinking about these counterexamples to Stalnaker's Thesis. For instance, we can prove:

Local Failures.
$Pr(\texttt{A} \rightarrow_{\ell}^{\mathbb{Z}} \texttt{B}) \neq P(\mathbf{B}|\mathbf{A})$ iff there is some $\mathbf{Z} \in \mathbb{Z}$ such that $P(\mathbf{Z}|\mathbf{A}) \neq P(\mathbf{Z}|\bar{\mathbf{A}})$.

Thus, failures of Restricted Agential Stalnaker's Thesis are predicted to arise only if (a) the conditional has a local interpretation and (b) the conditional is interpreted relative to a partition that is evidentially dependent on the conditional's antecedent.

However, since he does not supply a semantics that delivers contents for conditionals, Kaufmann's theory does not (by itself) account for the violations of Weak Sufficiency*. Furthermore, positing an ambiguity to account for the different interpretations of conditionals (and thus differing judgments about their probabilities) is actually unnecessary, once we have the local calculation. This is because if a local conditional is interpreted relative to the trivial partition $\{\mathbf{W}\}$, the conditional will be equivalent to its global cousin.[3] So, if you already have partition-dependent conditionals, there is no additional need for a non-partition-dependent conditional to account for the global interpretations: you can do so by allowing that the same, local conditional be interpreted relative to different partitions. Furthermore, Kaufmann's suggestion that the background partition is a chance-determining variable that is causally independent, but evidentially dependent, on the conditional's antecedent seems open to counterexample. Some violations of Restricted Coordinated Stalnaker's Thesis involve a background partition that does not mediate the objective chances of the consequent.[4] Our aim in what follows is to extract the lessons from Kaufmann's theory and use it to constrain the semantics

[3] Here is the quick calculation:

(i) $Pr(\mathtt{A} \rightarrow_{\ell}^{\{\mathbf{W}\}} \mathtt{B}) =$

(ii) $P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{Z}|\mathbf{W}) =$

(iii) $P(\mathbf{B}|\mathbf{A}) \cdot P(\mathbf{Z}) =$

(iv) $Pr(\mathtt{A} \rightarrow_{g} \mathtt{B})$

[4] Vann McGee's original case (McGee 2000) is like this. Here is a simplified version:

To tell the truth. You are watching a game show with three contestants, one of whom is a famous celebrity sworn to tell the truth, and the other of which are actors pretending to be the celebrity. Today's celebrity is the famous detective Sherlock Holmes. The three contestants are talking about a recent high-profile case involving the suspicious death of Rutherford Murdoch. You have convinced yourself that Murdoch's death was most likely an accident, although you admit that it is possible that he was killed by Brown, his business partner.

Based on the previous line of questioning, you are now almost positive that Contestant 1 is Holmes. Contestant 1 now says that he is convinced Murdoch was murdered and he is almost certain Brown was the killer. Either way, he says he is sure that:

(i) If Brown didn't kill Murdoch, someone else did.

for conditionals from Chapters 3 and 4 to predict the exceptional cases to our principles above.

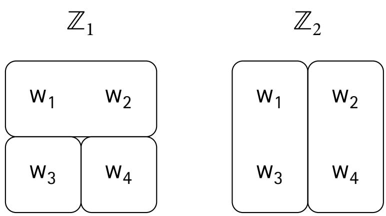
6.2 Partition Semantics

Here is our semantics so far:

> **Baseline Semantics**
> $[\![A \to B]\!]^{m,\sigma} = 1$ iff $\sigma(m_\sigma^{\mathbf{A}}) \in \mathbf{B}$

Notice that we have the modal base m provide a restriction on the worlds that matter to the evaluation of the conditional. Here, I propose that we generalize this constraint to the cell of a partition that the conditional is evaluated relative to.[5]

A partition $\mathbb{Z}$ on the set of possible worlds $\mathbf{W}$ is a set of disjoint subsets of $\mathbf{W}$ whose union is $\mathbf{W}$. Basically, $\mathbb{Z}$ divides up the worlds in $\mathbf{W}$ into n cells that share no members. So, where $\mathbf{W} = \{w_1, w_2, w_3, w_4\}$, both of the following are partitions of $\mathbf{W}$:



> On the basis of the testimony of someone whom you are almost sure is Holmes (who is sworn to tell the truth), you now think that (i) is very likely true. But now consider what would happen if you were to learn that Brown did not kill Murdoch. In that case, you would think it likely that Contestant 1 is not Holmes—after all, Contestant 1 was almost entirely convinced it was Brown, and the real Sherlock Holmes would not be so wrong in his assessment. But then you would discount the testimony in favor of (i) and return to your original view that, having excluded Brown as a possible murderer, it is most likely that Murdoch's death was an accident.

Analyzed within Kaufmann's theory, the relevant partition is whether Contestant 1 is Holmes. But the objective chance that someone else killed Murdoch given that Brown did not does not depend on whether Contestant 1 is Holmes or not.

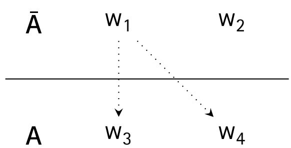
[5] As mentioned above, this move is also shared by Moss 2015, which in fact also argues, as I will later, that the partition is set by a question under discussion in the context. I return to compare our views below in the Appendix of this chapter.

Where $\mathbb{Z}$ is a partition on $\mathbf{W}$, let $w[\mathbb{Z}]$ be the cell of $\mathbb{Z}$ that contains w (similarly, let $\sigma[\mathbb{Z}] = \sigma_w[\mathbb{Z}]$). So, for instance, in the partitions above,

- $w_1[\mathbb{Z}_1] = \{w_1, w_2\}$
- $w_1[\mathbb{Z}_2] = \{w_1, w_3\}$

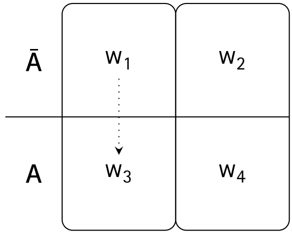
I propose that we add $w[\mathbb{Z}]$ as an additional constraint on the selected world for the conditional's semantics. Partitions provide world-dependent restrictions on conditional domains, thus demanding that we hold fixed similarity with respect to features picked out by the partition in evaluating a conditional. This forces us to compute the content of a conditional locally at each cell of the relevant partition, and then its overall content will just be the union of its content at each cell.[6]

We can illustrate how this works using an example. Suppose that $\mathbf{W} = \{w_1, w_2, w_3, w_4\}$ and that $\mathbf{A} = \{w_1, w_2\}$. Then, there are two possible "paths" from w_1 to an $\mathbf{A}$-world: one leading to w_3 and one leading to w_4:



These paths correspond to distinct sequences that share w_1 as their first member: $\sigma^1(\mathbf{A}) = w_3$ and $\sigma^2(\mathbf{A}) = w_4$.

Now, suppose we constrain the sequences to find the first $\mathbf{A}$-worlds that share the same partition cell, given partition $\mathbb{Z}_2 = \{\{w_1, w_3\}, \{w_2, w_4\}\}$:



[6] A similar thought is expressed by Stalnaker 2019: 194–195, using the apparatus of selection functions. Stalnaker proposes to generalize his Stalnaker 1975 "indicative constraint" on selection functions to partitions (see also Khoo 2016: 17).

Now, we can see that there is only one path from w_1 to an **A**-world that shares the same cell as w_1, namely w_3. Thus, it follows that $\sigma^1(\mathbf{A} \cap \sigma^1[\mathbb{Z}]) = \sigma^2(\mathbf{A} \cap \sigma^2[\mathbb{Z}]) = w_3$. Thus, partitions provide a way of coarsening refined contents, eliminating refined distinctions between possibilities.

We incorporate partition dependence into our semantics as follows:

Partitional Semantics
$[\![A \to B]\!]^{m,\mathbb{Z},\sigma} = 1$ iff $\sigma(m^{\mathbf{A}}_{\sigma} \cap \sigma[\mathbb{Z}]) \in \mathbf{B}$

As before, we define the content of a sentence by letting context initialize both the modal base and partition:

Content
$[\![A]\!]^c = \{\sigma\colon [\![A]\!]^{c,m_c,\mathbb{Z}_c,\sigma} = 1\}$

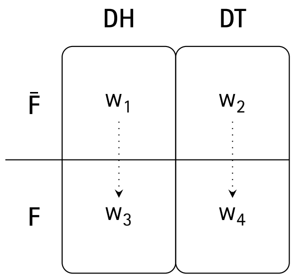
We will need to impose some additional constraints on the assignment of partitions to conditionals in context to ensure that this theory makes reasonable predictions. This is an issue we will come back to below. First, I want to point out how adopting Partitional Semantics will allow us to predict the exceptions to each of our generalizations above.

Recall Edgington's Coin:

Edgington's Coin. Smith has a coin, which is either double-headed or double-tailed (1/2 likely either way). We're not sure whether he flipped the coin, and whether he did or not is independent of whether it was double-headed or double-tailed.

(1) If Smith flipped the coin, it landed heads.

For now, let's assume that the relevant partition for the conditional is whether Smith's coin was double-headed or double-tailed, $\mathbb{D} = \{\mathbf{DH}, \mathbf{DT}\}$. We can illustrate the situation as follows:



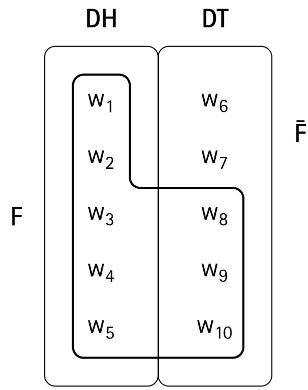
Notice, given this partition, that there is only one path from w_1 to an **F**-world, and only one path from w_2 to an **F**-world, as constrained by the partition $\mathbb{D}$. In effect, the partition forces us to hold fixed whether the coin is double-headed or double-tailed in evaluating what happens when it is hypothetically flipped. This means that the conditional $F \rightarrow H$ is equivalent, in the context, to that the coin is double-headed **DH**. And, since **DH** is logically stronger than $\mathbf{F} \supset \mathbf{H}$, we predict the exception to Weak Sufficiency*.

The same reasoning goes for McGee's Coins.

> **McGee's Coins**. McGee has 10 coins, each given to a different person, A, B, C, ... Five of the coins are double-headed and five are double-tailed. Each person knew the following: if they flip and it lands heads, they win \$1; if they flip and it lands tails, they lose \$.50; nothing happens if they don't flip. Three of the 5 people who received a doubled-headed coin knew their coin was double-headed; while 2 of the 5 people who received a double-tailed coin knew their coin was double-tailed; the rest thought they were flipping a fair coin. Everyone was rational and wanted to win.

(2) If A flipped, she won.

Assume again that the partition is whether A's coin was double-headed or double-tailed, $\mathbb{D} = \{\mathbf{DH}, \mathbf{DT}\}$. Then, we have:



Now, by Strong Centering, we have that $F \rightarrow W$ is true at all of the **DH**-worlds and false at the **DT**-worlds where A flipped (w_8, w_9, w_{10}). But now consider its truth value at the **DT**-worlds where A didn't flip (w_6, w_7). By Partitional Semantics, the truth of the conditional at these worlds depends on what happens at the selected **F**-world that shares their partitional cell. Thus, all such selected worlds will be in **DT** as well, and thus be worlds where A loses. So, $F \rightarrow W$ is false at all of the **DT**-worlds. Thus, its probability is equal to the probability of **DH**, namely 1/2.

But the probability of **W** given **F** is 5/8. We can see this by the fact that, of the eight worlds in which A flips, she wins at exactly five of them (the ones where the coin is double-headed).

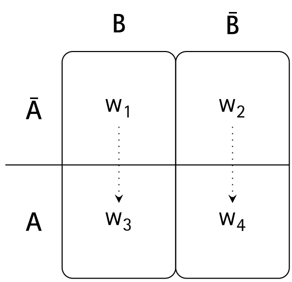
Generally, because of Local Failures, we also predict that conditionals will receive a "global" reading, whereby their probabilities are equal to the corresponding conditional probability, whenever the partition assigned to them is independent of its antecedent. The limiting case of this is when the conditional is assigned the trivial partition {**W**}. We will come back to these predictions below when we discuss how context determines what partition a conditional is evaluated relative to. Without constraints on the assignment of partitions to conditionals in context, the semantics threatens to wildly overgenerate readings of conditionals. More on this below.

Summary: ***Partition dependence***

Adopting a partition-constrained semantics of conditionals allows us to generate violations of Weak Sufficiency* and Restricted Agential Stalnaker's Thesis, in accordance with the kinds of examples where these principles seem to intuitively fail.

6.3 Partitional Pragmatics

The semantics from the previous section needs to be pragmatically constrained—otherwise, it threatens to predict not only that there will be trivial interpretations of conditionals but also that we should expect many more thesis-violating interpretations than we in fact find. To give just one example, suppose the partition for $A \to B$ is $\mathbb{B} = \{\mathbf{B}, \bar{\mathbf{B}}\}$, where both **AB** and $\mathbf{A\bar{B}}$ are epistemically possible. Then, we would predict that $A \to B$ is epistemically equivalent to B—that is, we predict they are true at the same epistemically possible worlds. To see why, consider the following diagram:



Since every world at the **B**-cell must be mapped to an **AB**-world, the conditional will be true at each **B**-world in the information state (w_1, w_3 above). And since every world at the $\bar{\mathbf{B}}$-cell must be mapped to an $\mathbf{A\bar{B}}$-world, the conditional will be false at each $\bar{\mathbf{B}}$-world in the information state (w_2, w_4 above). Thus, it follows that $A \to B$ and B

will be epistemically equivalent when the conditional is assigned the partition $\{\mathbf{B}, \bar{\mathbf{B}}\}$.

This result is clearly quite bad. Conditionals do not seem to have interpretations on which they are equivalent to their consequents.[7] So, we must have some way of pragmatically ruling out such interpretations, even though they are allowed by the semantics. In this section, I motivate constraints on the assignment of partitions to conditionals that will rule out degenerate interpretations like the one above, and will also predict why conditionals that violate our generalizations are rare.

My proposal is that the relevant background partition in a context will often be a salient question under discussion. I borrow the notion of a question under discussion (QUD) from Craige Roberts (Roberts 1996, 2012b, a). Roberts holds that contexts determine a stack of QUDs, each of which provides a discourse goal—answering that question. It is natural to think that a question determines a partition on the epistemically possible worlds in the context (cf. Groenendijk & Stokhof 1984); the discourse goal of answering a question is to jointly determine which cell of the partition we are in. With that in mind, thinking that questions under discussion are the partitions conditionals are interpreted relative to, is a reasonable thought. Furthermore, following Roberts, we allow that questions may be implicitly under discussion—meaning that a question need not have been explicitly asked to be under discussion; instead, it may be revealed to be under discussion by the fact that other speech acts are understood to be relevant to answering it. Roberts also allows that multiple questions may be under discussion in one context, but they are arranged in a stack, ordered by salience. Finally, notice that, in addition to each of the QUDs, there will always be one partition that is minimally salient in the context, in the sense that it is always there, and requires no effort to draw attention to it: the trivial partition $\{\mathbf{W}\}$. Since it is always around, it is always less salient than any QUD in the context. Nonetheless, it is a possible partition a conditional may be assigned in some contexts.

[7] A related worry is raised by Douven 2008.

The two pragmatic constraints on the partitions assigned to conditionals I will motivate are:

> **Pragmatic Constraints on Partitions**
> $\mathbb{Z}$ can be the partition for $A \rightarrow B$ in c only if it is common ground at c that:
>
> a. $\forall \mathbf{Z} \in \mathbb{Z} : \mathbf{Z} \cap m_c^{\mathbf{A}} \neq \varnothing$, and
> b. $\exists \mathbf{Z} \in \mathbb{Z} : \mathbf{Z} \nsubseteq \mathbf{B}$ and $\mathbf{Z} \nsubseteq \bar{\mathbf{B}}$.

The first condition is that each cell of $\mathbb{Z}$ contains at least one **A**-world in the conditional's modal base. The second condition ensures that **B**/$\bar{\mathbf{B}}$ are suitably independent of $\mathbb{Z}$. Together, these conditions will rule out interpretations of $A \rightarrow B$ on which it is equivalent to B.

Recall that, according to Partitional Semantics, $[\![A \rightarrow B]\!]^{m,\mathbb{Z}}$ is true at a sequence σ only if the first $m_\sigma^{\mathbf{A}}$-world in $\sigma[\mathbb{Z}]$ is a **B**-world. Thus, if $\sigma[\mathbb{Z}]$ contains no $m_\sigma^{\mathbf{A}}$-worlds, the semantics would crash—the conditional is undefined at such a sequence. Thus, Partitional Semantics motivates the first definedness constraint on background partitions.

The second constraint is motivated by considerations of speech economy. Given the (idealized) aim of communicating the maximum amount of information with the least amount of effort, we expect speakers not to use a complex sentence S to (knowably) convey the same amount of information that they could have otherwise conveyed by using a sentence S' that is a proper part of S. Versions of such a constraint have been proposed to account for diachronic linguistic change and pragmatic interpretation (cf. Horn 1993 and references therein). Since hearers expect speakers to be economical, it is natural that speaker interpretation would normally be guided by a principle like the following:

> **Avoid Trivialization**
>
> Interpret a complex sentence so that it is not epistemically equivalent with any of its sentential constituents.

Given a principle like Avoid Trivialization, we expect two things. The first is that, when there is no way to interpret a complex sentence except as equivalent to one of its sentential constituents, such a sentence should be defective (we expect economical speakers would then just assert the constituent, which means the same and is less complicated). The second is that we should not find interpretations of a complex sentence like $O(A)$ to be equivalent to A.

An instance of the first prediction is the observation known as Hurford's Constraint (Hurford 1974). Disjunctions for which one disjunct entails the other are generally felt to be quite odd:

(3) a. #John is from Paris, or France.
    b. #John is not married, or is a bachelor.

A natural explanation appeals to a principle like Avoid Trivialization—since these sentences cannot be interpreted in accordance with the principle, they are predicted to be odd (cf. Katzir & Singh 2013).

To see an instance of the second prediction, take the possibility modal *could*. Modals like this exhibit underspecification—they can be used to express epistemic possibility, metaphysical possibility, or deontic possibility:

(4) a. Smith could be in Athens. Epistemic
    b. Smith could go to Athens (nothing is hold him back). Metaphysical
    c. Smith could refrain from saving the child (he is not required to help). Deontic

On the standard Kratzerian semantics for modals, this underspecification is resolved by pairing the modal with an appropriate modal base, which maps a world to a set of worlds that the modal quantifies over. For Kratzer, this pairing happens in context, perhaps subject to lexical and pragmatic constraints.[8] One such pragmatic constraint seems to be **Avoid Trivialization**. Notice that, without it, we would be able to interpret *could* relative to a modal base m that maps each world w

[8] See Condoravdi 2002 and Chapter 7 for more on some possible lexical constraints on the assignment of modal domains to modals in context.

to the singleton $\{w\}$. If it did, then *could A* and *A* would be necessarily equivalent.

$$\llbracket could\ A \rrbracket^{w,m} = 1 \text{ iff}$$
$$\exists w' \in m_w\text{: } \llbracket A \rrbracket^{w',m} = 1 \text{ iff}$$
$$\exists w' \in \{w\} : \llbracket A \rrbracket^{w',m} = 1 \text{ iff}$$
$$\llbracket A \rrbracket^{w,m} = 1$$

However, there does not seem to be a plausible interpretation of *could A* on which it is necessarily equivalent to *A*. Take (4-a), for instance: I, at least, cannot hear any interpretation of (4-a) on which it is necessarily equivalent to *Smith is in Athens*. The lack of any such interpretation of possibility modals is evidence then for Avoid Trivialization, which predicts this. Furthermore, for the same reason, given this principle, we can immediately rule out partitions (like $\{\mathbf{B}, \bar{\mathbf{B}}\}$) since such a partition would make $A \rightarrow B$ epistemically equivalent to *B*. Thus, Avoid Trivialization motivates our second pragmatic constraint above.[9]

[9] We might wonder whether discourses involving explicit reasoning raise trouble for Avoid Trivialization. Consider the following piece of Modus Ponens reasoning:

(i) A: It is raining.
B: Yes, and if it is raining, John brought an umbrella. Therefore, John brought an umbrella.

Notice that, when speaker B asserts the conditional, it is at that point common ground between A and B that it is raining. However, given Strong Centering, it follows that the conditional *if it is raining, John brought an umbrella* is equivalent throughout the common ground with its consequent *John brought an umbrella*. Cases like these show that Avoid Trivialization can be overruled in some circumstances; the reason in this case is that, in uttering the conditional, the speaker commits herself to an epistemic connection between rain and John bringing an umbrella. Since her commitment to this epistemic connection plausibly does not depend on her commitment to it raining, it is plausibly informative enough in the context to warrant assertion, which is why I think it is not felt to be infelicitous. In the Paris or France example (3-a) above, it is just much harder to grasp what additional commitment the speaker might be trying to communicate that would justify the extra disjunct (*or Paris*). However, notice that if the speaker instead had said,

(ii) John is from Paris, or anyway somewhere in France.

This is felicitous, and I think it is because it communicates something about the speaker's relative levels of confidence: in particular, that she's ultimately committed to John being from France, and secondarily committed to Paris being the most likely place in France where he's from.

6.3.1 Applying the Partitional Constraint

Start with McGee's Coins—the same points will apply to Edgington's Coin as well. Recall:

> **McGee's Coins**. McGee has 10 coins, each given to a different person, A, B, C, ... Five of the coins are double-headed and five are double-tailed. Each person knew the following: if they flip and it lands heads, they win \$1; if they flip and it lands tails, they lose \$.50; nothing happens if they don't flip. Three of the 5 people who received a doubled-headed coin knew their coin was double-headed; while 2 of the 5 people who received a double-tailed coin knew their coin was double-tailed; the rest thought they were flipping a fair coin. Everyone was rational and wanted to win.

(2) If A flipped, she won.

Here, there are several salient questions under discussion:

- Whether A flipped the coin: $\mathbb{F} = \{\mathbf{F}, \bar{\mathbf{F}}\}$
- Whether A won or didn't win: $\mathbb{W} = \{\mathbf{W}, \bar{\mathbf{W}}\}$
- Whether A's coin was double-headed or double-tailed: $\mathbb{D} = \{\mathbf{DH}, \mathbf{DT}\}$
- Whether A knew her coin was double-headed or double-tailed: $\mathbb{K} = \{\mathbf{K}, \bar{\mathbf{K}}\}$

Given our pragmatic constraint, we can immediately rule out $\mathbb{F}$ and $\mathbb{W}$:

> **Pragmatic Constraints on Partitions**
> $\mathbb{Z}$ can be the partition for $A \to B$ in c only if it is common ground at c that:
>
> a. $\forall \mathbf{Z} \in \mathbb{Z} : \mathbf{Z} \cap m_c^{\mathbf{A}} \neq \emptyset$, and
> b. $\exists \mathbf{Z} \in \mathbb{Z} : \mathbf{Z} \not\subseteq \mathbf{B}$ and $\mathbf{Z} \not\subseteq \bar{\mathbf{B}}$.

And we can also rule out $\mathbb{K}$ since we have stipulated that the individuals are rational and want to win; thus, if A knew her coin was double-tailed, she would not flip, and thus there are no worlds where A's coin is double-tailed, she knows this, and she flips. But then there is a cell of $\mathbb{K}$ that contains no **F**-worlds.

This leaves us with whether A's coin was double-headed or double-tailed, $\mathbb{D}$, and thus, since this question is salient in the context, we predict that this is the partition the conditional is evaluated relative to.

Turn next to the Coin Flip case from footnote 2. Recall that, in the setup of that case, you think that the coin is weighted 3 to 1 in favor of tails, so your unconditional probability in tails is $P(\mathbf{T}) = 3/4$. You also think that it is $9/10$ likely that Jones said "heads" so $P(\mathbf{S_H}) = 9/10$.

Recall the two conditionals under consideration:

(i) If Jones is correct, the coin landed heads. $C \rightarrow H$

(ii) If Jones is correct, she said that the coin landed heads. $C \rightarrow S_H$

Finally, in the context as described, there are three salient questions under discussion:

- Whether the coin landed heads or tails: $\mathbb{H} = \{\mathbf{H}, \mathbf{T}\}$
- Whether Jones said "heads" or "tails": $\mathbb{SH} = \{\mathbf{S_H}, \mathbf{S_T}\}$
- Whether Jones is correct: $\mathbb{C} = \{\mathbf{C}, \bar{\mathbf{C}}\}$.

For each of (i)/(ii), exactly one of these partitions meets the pragmatic constraint on partitions. Start with $C \rightarrow H$. Neither $\mathbb{C}$ nor $\mathbb{H}$ are admissible partitions for this conditional. $\mathbb{C}$ is ruled out by the first constraint, since $\bar{\mathbf{C}}$ is not epistemically compatible with its antecedent, **C**. And $\mathbb{H}$ is ruled out by the second constraint, because it would trivialize the conditional. In contrast to both, $\mathbb{SH}$ is admissible: both of its cells are compatible with **C** and this partition does not trivialize the conditional. Thus, barring other interpretive pressures, the most salient partition for $C \rightarrow H$ meeting the pragmatic constraints will be $\mathbb{SH}$ in this context.

Next, consider $C \rightarrow S_H$. Neither $\mathbb{C}$ nor $\mathbb{S}_{\mathbb{H}}$ are admissible partitions for this conditional. $\mathbb{C}$ is ruled out by the first constraint, since $\bar{\mathbf{C}}$ is not epistemically compatible with its antecedent, **C**. And $\mathbb{SH}$ is ruled out by the second constraint, because it would trivialize the conditional. In contrast to both, $\mathbb{H}$ is admissible: both of its cells are compatible with

C and this partition does not trivialize the conditional. Thus, barring other interpretive pressures, the most salient partition for $C \rightarrow S_H$ meeting the pragmatic constraints will be $\mathbb{H}$ in this context.

Given these predictions, we can show that $Pr(\mathtt{C} \rightarrow_m^{\mathbb{SH}} \mathtt{H}) \neq Pr(\mathtt{C} \rightarrow_m^{\mathbb{H}} \mathtt{S_H})$. Here is the calculation:

$$\begin{aligned}
&Pr(\mathtt{C} \rightarrow_m^{\mathbb{SH}} \mathtt{H})\\
&\qquad = P(\mathbf{H}|\mathbf{CS_H}) \cdot P(\mathbf{S_H}) + P(\mathbf{H}|\mathbf{CS_T}) \cdot P(\mathbf{S_T})\\
&\qquad = 1 \cdot P(\mathbf{S_H}) + 0 \cdot P(\mathbf{S_T})\\
&\qquad = P(\mathbf{S_H})
\end{aligned}$$

$$\begin{aligned}
&Pr(\mathtt{C} \rightarrow_m^{\mathbb{H}} \mathtt{S_H})\\
&\qquad = P(\mathbf{S_H}|\mathbf{CH}) \cdot P(\mathbf{H}) + P(\mathbf{S_H}|\mathbf{CT}) \cdot P(\mathbf{T})\\
&\qquad = 1 \cdot P(\mathbf{H}) + 0 \cdot P(\mathbf{T})\\
&\qquad = P(\mathbf{H})
\end{aligned}$$

Furthermore, these predicted values match our intuitions in the case: $P(\mathbf{S_H})$ is high, since you think it antecedently likely that Jones said "heads," and thus (i) is predicted to be likely; but $P(\mathbf{H})$ is low, since you think it antecedently likely that the coin landed on tails, and thus (ii) is predicted to be unlikely.

6.3.2 Overriding the Default

We now come to a complication: intuitions in cases like McGee's Coin and Coin Flip are not uniform. Some people report thesis-violating intuitions, while others report thesis-conforming intuitions. Consider McGee's Coin, and imagine someone reasoning as follows:

(5) Suppose A flipped the coin. Then it is likely the coin was double-headed, since if it were double-tailed and he knew this, he wouldn't have flipped it.

(6) But if Smith flipped it and it was double-headed, then it landed heads and he won.

(7) So, it is likely that if Smith flipped the coin, he won.

Such reasoning is plausible, and it results in the conclusion that (2) is thesis-conforming. How can we reconcile this with our partition-dependent semantics? My strategy draws on a principle of charity. In cases like these where someone proposes an argument for some conclusion, there is pressure from charity to interpret the argument in such a way that it is not equivocal (if such an interpretation is possible, given other pragmatic considerations). However, in the argument above, the only way to avoid interpreting the premises and conclusion equivocally is to assign each conditional the trivial partition $\{\mathbf{W}\}$. And, we know that if a (simple) conditional is interpreted relative to the trivial partition, it will be thesis-conforming. That is the sketch of my strategy here. Let us now explore the details a bit more carefully.

In the context of McGee's Coin, recall that there are four questions under discussion:

- Whether A flipped the coin: $\mathbb{F} = \{\mathbf{F}, \bar{\mathbf{F}}\}$
- Whether A won or didn't win: $\mathbb{W} = \{\mathbf{W}, \bar{\mathbf{W}}\}$
- Whether A's coin was double-headed or double-tailed: $\mathbb{D} = \{\mathbf{DH}, \mathbf{DT}\}$
- Whether A knew her coin was double-headed or double-tailed: $\mathbb{K} = \{\mathbf{K}, \bar{\mathbf{K}}\}$

However, none of these supply a partition that is admissible by the Pragmatic Constraint for both (6) and (7). We have already seen that only $\mathbb{D}$ is admissible for (7). And this partition is not admissible for (6) since one of its cells, **DT**, is not compatible with the antecedent of (6).

Thus, given that there are no other salient questions in the context, the only partition meeting the Pragmatic Constraint for both (6) and (7) is the trivial partition $\{\mathbf{W}\}$. Given the interpretive pressure to interpret the conditionals (6) and (7) as non-equivocal, this means that they must be interpreted relative to the same partition, and thus (6)/(7) must also be interpreted relative to the trivial partition. Thus, we predict that, in such a context, (7) will receive a thesis-conforming interpretation.[10]

Summary: *Partitional pragmatics*

Combining our partition-constrained semantics with a suitable pragmatic principle governing the assignment of partitions to conditionals in context allows us to confine the violations of Weak Sufficiency* and Restricted Agential Stalnaker's Thesis as exceptional cases.

6.4 Conclusion

We have now reached the conclusion of Part II on the probabilities of conditionals. We've seen how to adapt a sequence semantics to capture the probabilities of conditionals while avoiding well-known triviality results; and we've also seen how to refine our semantics by introducing partition dependence to predict exceptional cases that violate Weak Sufficiency* and Restricted Agential Stalnaker's Thesis. In Part III, we turn from indicative conditionals to subjunctive conditionals, starting with an account of why their morphological differences give rise to their semantic differences.

Appendix: Comparison with Moss

Moss 2015, 2018 develops a view similar to the one developed here. In particular, Moss proposes that conditionals and probability operators are evaluated relative to partitions, and her theory is able to account for the Thesis-violating intuitions discussed so far. However, Moss's theory is in many respects quite different from mine. For one, Moss

[10] See Khoo 2016 for more discussion and additional examples.

proposes partition sensitivity for a wide range of expressions, including conditionals, but also logical operators and modals. She uses this to capture embeddings of non-propositional contents in modal and logical environments. As a result, when a probability operator embeds a conditional, which is non-propositional for her (denoting a set of credal spaces), it ends up operating on a proposition—the union of propositions that "accept" the conditional relative to the probability operator's partition.

I criticized this feature of Moss's theory in Chapter 2, arguing that as a result, her theory faces a change for predicting the standard, Thesis-conforming probabilities of conditional contents. However, I also noted there that Moss's theory can correctly predict Thesis-conforming intuitions for low-scope probabilified conditionals—that is, that (8) is true iff the probability that Smith got heads given that he flipped the coin is greater than 1/2:

(8) If Smith flipped the coin, he probably got heads.

To handle Thesis-conforming intuitions about high-scope probabilified claims (9), or our judgments about the probabilities of bare conditional sentences,

(9) Probably if Smith flipped the coin, he got heads.

I suggested that Moss could instead offer an error theory whereby someone with those intuitions is really judging the truth of a related low-scope probabilified conditional, or perhaps she could adopt one of the other error theories discussed in that chapter.

Here, I want to show how Moss's theory does nicely capture Thesis-violating intuitions about the probabilities of conditionals, but cannot capture the variability of Thesis-violating intuitions. This point undermines the potential error theory Moss might attempt to use to account for Thesis-conforming intuitions, since we sometimes have both Thesis-violating and Thesis-conforming intuitions about one and the same wide-scope probability claim about a conditional (or judgment about the probability of a conditional).

Let's start by seeing how Moss's theory captures the Thesis-violating intuitions by focusing on a variation of a case she gives:

> **Jump**. Jill was one of three people poised to jump from their respective buildings. You know that two of the three people had nets under them, and two of the three of them jumped. You know that one of the jumpers lived, that jumping without a net results in death, while jumping with a net results in survival. Thus, you know that one of the two who jumped did so with a net underneath them.

Schematically, we have three equally likely possible worlds: w_1, w_2, w_3, where:

- **J** = that Jill jumped = $\{w_2, w_3\}$
- **N** = that there was a net under Jill = $\{w_1, w_2\}$
- **D** = that Jill died = $\{w_3\}$

Consider: What value of *n* yields a true statement from (10)?

(10) It is *n* probable that if Jill jumped, she died.

One natural intuition is 1/3—this is the Thesis-violating judgment brought out by the following reasoning: *the probability that there was no net underneath Jill is* 1/3*; she lived if she jumped with a net and died if she jumped without one; so it's* 1/3 *probable that if Jill jumped, she died*. But another natural intuition is 1/2—this is the Thesis-conforming judgment brought out by the reasoning: *given that Jill jumped, the probability that there was a net under her is* 1/2*; thus, it is* 1/2 *probable that if Jill jumped, she died*.

Moss's theory nicely captures the Thesis-violating intuition. Here, again, is the statement of Moss's semantics for *n-probably* and $A \rightarrow B$ (see Moss 2015: 31–34, Moss 2018: 232–235):

$$[\![\text{Probably}_i^{=n}(A)]\!]^{g,m} = 1 \text{ iff } m(\bigcup\{p \in g(i) \colon [\![A]\!]^{g,m^p} = 1\}) = n$$

$$[\![A \rightarrow_j B]\!]^{g,m} = 1 \text{ iff}$$

$$m(\bigcup\{q \in g(j) : m^q(\mathbf{B}) = 1\} | \bigcup\{p \in g(j) : m^p(\mathbf{A}) = 1\}) = 1$$

Take the sentence and suppose we're in a context in which $g(i) = \{\mathbf{N}, \bar{\mathbf{N}}\}$ and $g(j) = \{\mathbf{J}, \bar{\mathbf{J}}\}$ (in line with Moss 2018: 16–19):

(11) It is 1/3 probable that if Jill jumped, she died.

$$\text{Probably}_i^{=1/3}(J \rightarrow_j D)$$

We thus compute that (11) is true:

- $[\![\text{Probably}_i^{=1/3}(J \rightarrow_j D)]\!]^{g,m} = 1$ iff
  (*) $m(\bigcup\{p \in g(i): [\![J \rightarrow_j D)]\!]^{g,m^p} = 1\}) = 1/3$
- Then, since $g(i) = \{\mathbf{N}, \bar{\mathbf{N}}\}$ we must check whether $[\![J \rightarrow_j D]\!]^{g,m^{\mathbf{N}}} = 1$ and $[\![J \rightarrow_j D]\!]^{g,m^{\bar{\mathbf{N}}}} = 1$
  - $[\![J \rightarrow_j D]\!]^{g,m^{\mathbf{N}}} = 1$ iff
    $m(\bigcup\{q \in g(j) : m^{\mathbf{N} \wedge q}(\mathbf{D}) = 1\} | \bigcup\{p \in g(j) : m^{\mathbf{N} \wedge p}(\mathbf{J}) = 1\}) = 1$
    And this does not hold because $m^{\mathbf{N} \wedge \mathbf{J}}(\mathbf{D}) \neq 1$, and $m(\varnothing | \mathbf{J}) = 0$.
  - $[\![J \rightarrow_j D]\!]^{g,m^{\bar{\mathbf{N}}}} = 1$ iff
    $m(\bigcup\{q \in g(j) : m^{\bar{\mathbf{N}} \wedge q}(\mathbf{D}) = 1\} | \bigcup\{p \in g(j) : m^{\bar{\mathbf{N}} \wedge p}(\mathbf{J}) = 1\}) = 1$
    And this holds because $m^{\bar{\mathbf{N}} \wedge \mathbf{J}}(\mathbf{D}) = 1$, and $m(\mathbf{J}|\mathbf{J}) = 1$.
- Therefore, (*) holds iff $m(\bar{\mathbf{N}}) = 1/3$. And it does!

However, for the reasons discussed at the end of Chapter 2, Moss's theory cannot predict the Thesis-conforming intuition, on which (12) is true:

(12) It is 1/2 probable that if Jill jumped, she died.

Again, the reason is that there are no obvious partitions that are salient which are such that the probability of the union of some subset of their cells is 1/2. And the reason is the same as with Hájek 2011a's "wallflower proof": all of the salient partition cells (any Boolean combination of **J**, **N**, **D** and their negations) have a probability that is a multiple of 1/3.

To bring out the contrast between our theories, notice that my theory predicts both Thesis-violating and Thesis-conforming intuitions in this kind of case. The Thesis-violating intuition is the result of the embedded conditional being interpreted relative to the net partition, while the Thesis-conforming intuition is the result of it being interpreted relative to the trivial partition.

PART III

SUBJUNCTIVE VS. INDICATIVE

7
Subjunctive Conditionals: The Role of Tense

In this part of the book, we will switch our focus to the relationship between indicative conditionals like (1):

(1) If John is not in Athens, he is in Barcelona.

and subjunctive conditionals like (2):

(2) If John had not been in Athens, he would have been in Barcelona.

Indicative and subjunctive conditionals differ in at least three ways. One difference is **morphological**: (2) contains the past perfect *had* in its antecedent and the auxiliary *would* and present perfect *have* in its consequent, while (1) lacks all three. A second diffference is **semantic**—(1) is about how things are, given what we know (plus some hypothetical information), while (2) is about alternative courses of history. Knowing merely that John is either in Athens or in Barcelona (without knowing which) is sufficient for knowing (1) (cf. our discussion of Or-to-If in Chapters 1–3), but this is not sufficient for knowing (2). Knowing the latter requires knowing something about John's backup travel plans, or where he was flying to and from, or some other causally relevant information. Finally, the two conditionals differ **pragmatically**—utterances of (1) typically signal that the speaker thinks it is epistemically possible that John is not in Athens, while utterances of (2) typically signal that the speaker thinks this is not epistemically possible.

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0008

With these distinctions in hand, I will use *indicative*/*subjunctive* to mark the first, morphological, difference.[1] Indicative conditionals are those that lack the extra tense morphology found on subjunctive conditionals, tense morphology that is not doing what it normally does. Notice, for instance, that there need not be any difference in the times the antecedents/consequents of (1) and (2) are about; we can insert *right now* in each:

(3) a. If John is not in Athens right now, he is in Barcelona (right now).
b. If John had not been in Athens right now, he would have been in Barcelona (right now).

However, when it appears outside of a conditional, *had* locates the time of the described event as before some past reference time:

(4) The plane had left before I arrived.

In (4), we know that the plane's leaving was completed before my past arriving. To see that past perfects locate events in the past, notice that the following is defective:

(5) ??The plane had left tomorrow.

Things are different when we embed this sentence under an *if*-clause. Here is an example, embedded in a broader discourse to bring out the intended reading:

(6) I thought the flight was tomorrow, not today! If the plane had left tomorrow, I would have caught it.

[1] As a brief aside: the labels *indicative* and *subjunctive* are problematic, since the distinction is not one of mood (see Iatridou 2000, 2014). Despite these misgivings, since the terminology is rather entrenched at this point, for the purposes of this book, treat my use of these terms to track the morphological differences described here.

Unlike (5), (6) is totally acceptable: we know exactly what it means. Thus, when embedded under *if*, *had* seems to lack the temporal meaning it has when unembedded. Following Iatridou 2000, we will call it a **fake tense.**[2] So, the morphological difference amounts to this: subjunctive conditionals contain fake past tense, while indicatives lack fake past tense. Research in morphosemantics shows that this difference is robustly cross-linguistic—that is, languages use the same kind of morphology to mark the same semantic/pragmatic differences between conditionals like (1) and (2).[3]

In addition, following Ippolito 2003, we can distinguish two-past subjunctives like (2) from one-past subjunctives, which lack the perfects in both antecedent and consequent:[4]

(7) If John were not in Athens, he would be in Barcelona.

Next, I will use *epistemic/metaphysical* to mark the second, semantic difference. (1) is epistemic, in the sense that it is about our current evidence. By contrast, (2) is not epistemic; rather, it is about how the history of the world would have proceeded under certain supposed changes (hence the label *metaphysical*). This difference is clearly felt in Adams's classic minimal pair (from Adams 1965):

(8) If Oswald did not shoot Kennedy, someone else did.

(9) If Oswald had not shot Kennedy, someone else would have.

[2] Iatridou 2000 also identifies fake aspect on the translations of conditionals like (2) in many other languages. These involve imperfective morphology (roughly, this is realized as progressive aspect in English) that does not have its usual imperfective meaning.

[3] See Iatridou 2000: 245, 263–266; von Fintel & Iatridou 2008: 120–126; and also Steele 1975; James 1982. This work surveys Romance and Germanic languages, as well as Modern Greek, Papago, Proto-Uto-Aztecan, Japanese, Korean, Hebrew, Turkish, and Basque. Other languages, such as Hungarian, use a dedicated morpheme for subjunctive conditionals.

[4] In fact, the crucial bit seems to be the missing present perfect in the consequent. Thus, the following should count as a two-past subjunctive:

(i) If John were not in Athens, he would have been in Barcelona.

The fact that we know Kennedy was shot (and thus shot by Oswald or someone else) is sufficient for the truth of (8) (in our mouths). But this fact is not sufficient for the truth of (9). Rather, (9) conveys that Kennedy's being shot was in some sense inevitable—that he would have been shot, either by Oswald or by someone else. Since this is not the case, (9) is false. So, the semantic difference is a difference in truth conditions for these two conditionals.

Finally, I will use *open/counterfactual* to mark the third, pragmatic difference. As used on a particular occasion, a conditional *if A, B* is open iff the speaker on that occasion conveys that they think *A* is epistemically possible. And, as used on a particular occasion, a conditional *if A, B* is counterfactual iff the speaker on that occasion conveys that they think *A* is false. Typically, (uses of) subjunctives are counterfactual, while (uses of) indicatives are open (compare (1) and (2) above).[5] However, there are non-counterfactual subjunctives:

(10) If Jones had taken arsenic, he would have shown just exactly those symptoms that he does in fact show. (Anderson 1951)

(11) If my plants died tomorrow, I would not be happy. (Arregui 2009)

As Anderson notes, (40) "would probably be taken as lending support to the view that Jones took arsenic—it would certainly not be held to imply that Jones did not take arsenic" (Anderson 1951: 37). And as Arregui notes, (11) is infelicitous if it is presupposed that the plants have already died (and thus cannot die tomorrow).

[5] Indeed, it seems that all uses of indicatives are open. It might seem that an exception is from conditionals uttered as epistemic backups:

(i) A: Oswald killed Kennedy.
B: That is not a settled fact.
A: Um, OK. Well at least grant me this: if Oswald did not shoot Kennedy, someone else did.

However, we may still hold that backup uses of conditionals are open—the relevant epistemic possibilities are those that are not ruled out by the evidence both speaker and hearer agree about.

These morphological, semantic, and pragmatic differences are clearly related. What we want is a theory that explains why the morphological differences give rise to the semantic and pragmatic ones. Broadly, there are two kinds of explanations on offer. The first approach treats the meaning of the past tense morphology on subjunctive conditionals as modal distancing rather than temporal distancing. Theories that take up this strategy are called **modal past** theories.[6] Thus, on modal past theories, the domain of the subjunctive conditional is interpreted to extend beyond the epistemically possible worlds. The second kind of explanation treats the meaning of the past tense morphology on subjunctive conditionals as temporal shifting that manipulates a parameter of the conditional itself rather than affecting its antecedent or consequent. Theories that take up this strategy are called **temporal past** theories.[7] On temporal past theories, a subjunctive conditional like (2) just is the past-tensed version of a future-directed indicative conditional, (12):

(12) If John is not in Athens, he will be in Barcelona.

Such views must then appeal to the interaction of this past-shifting with additional principles to account for the semantic and pragmatic differences between indicative and subjunctive conditionals.

In the remainder of this chapter, I will critically discuss some modal past theories. As we will see, I am skeptical that such theories can be made to work. Still, temporal past theories face their own challenges, too, for it is not obvious how the mere presence of past tense morphology on an indicative conditional could lead to the semantic differences between (for example) (1) and (2). In Chapter 8, I present my favored version of a temporal past theory, and argue that it provides a plausible explanation of why the distinctive morphological features of subjunctive conditionals give rise to their distinctive semantic and pragmatic properties.

[6] The most prominent recent defender of this view is Iatridou 2000, which draws inspiration from Isard 1974; Lyons 1977. More recently, the view has been advocated by Starr 2013; Schulz 2014.

[7] This kind of view has been defended by Tedeschi 1981; Dudman 1983, 1984, 1988; Edgington 1995; Ippolito 2003, 2006, 2013; Arregui 2005, 2007, 2009; Khoo 2015.

7.1 Against the Modal Past

We may divide modal past theories into two camps (cf. Iatridou 2000): those which hold that the modal meaning of the fake tense in subjunctive conditionals results from an accidental homophony, and those which hold that it results from the fact that the meaning of tense morphology is underdetermined and can change in different linguistic environments. I will set aside views in the first camp as a last resort, since, ideally, we expect that such a cross-linguistically robust generalization would not be the result of an accidental homophony. Here, I will focus on views in the second camp. Such views must explain:

(a) What the thin meaning of past tense morphology is, and
(b) How this thin meaning combines with various linguistic environments to result in the temporal past tense meaning and modal past tense meaning.

Iatridou 2000 presents the best-known modal past theory. According to her theory, the thin meaning of past tense just is:

> $T(x)$ excludes $C(x)$

Here, '$T(x)$' stands for the topic x—this is the x that we are talking about (or quantifying over)—and '$C(x)$' stands for the x that for all we know is the x of the speaker. When not embedded in an *if*-clause or a desire report (as in *I wish I were rich*), x ranges over times. So, in such environments, the meaning of past tense just is its ordinary, temporal meaning:

> Topic time excludes the time of utterance (as far as we know).

By contrast, Iatridou holds that when embedded in an *if*-clause or desire report, the domain of x is a set of worlds. In such environments, the meaning of past tense just is:

Topic world(s) excludes the world of utterance (as far as we know).

Iatridou's account has the potential to account for the semantic and pragmatic differences between indicative and subjunctive conditionals. Regarding the semantic difference, if we suppose that the default topic worlds are the worlds that, for all we know, may be actual (just as the default topic time is the time of utterance, or the times that, for all we know, may be the utterance time), then we expect indicative conditionals, which lack a fake past, to be interpreted as quantifying over the epistemically possible worlds. But if the non-epistemically possible worlds are metaphysically possible, then we expect subjunctive conditionals, which quantify over worlds that exclude the epistemically possible, to quantify over metaphysically possible worlds that are not epistemically possible. This difference in modal flavor may account for the semantic difference we articulated above.

Regarding the pragmatic difference, given that indicatives quantify over epistemically possible worlds, there must be some reason why a speaker would choose to use a subjunctive conditional to talk about non-epistemically possible worlds. The most natural reason is that she thinks the antecedent or consequent of that conditional is not epistemically possible. Hence, the view has the resources to predict that uses of subjunctive conditionals are typically counterfactual. Interestingly, the view does not predict that all uses are counterfactual—it is compatible with the fact that every non-epistemically possible A-world is a B-world that there are epistemically possible A-worlds—so a speaker may cancel this counterfactuality implicature without contradicting herself, as presumably happens with the Anderson example.

Despite these nice predictions, Iatridou's theory faces several challenges.

For one, the view is too open-ended, as it stands. In aiming for a simple parallel between the modal and temporal distancing, Iatridou ends up predicting that, in the temporal case, past tense just means that the topic time is outside of the utterance time. But what we want is something more specific: that topic time *precedes* utterance time. This worry can be put as a question that Iatridou's theory leaves unaddressed, which is why we do not find temporal interpretations of past tense morphology that places the topic time to the future of utterance time.[8] Compounding this worry is the fact that there are other dimensions of evaluation that have default values but can be excluded, such as the location of the utterance (*In Paris*, *Wherever Smith goes*, etc.), or the identity of the speaker (*As for you*, monsters in Amharic; cf. Schlenker 2003). Given the schematic meaning for past tense, we would expect there to be possible interpretations on which location or speaker identity are excluded in the same way time and world are. Yet, I have found no instances of past tense morphology that do this. My sense is that the solution to this challenge may come from the compositional derivation of temporal past tense meaning, but the solution space is not obvious.

Second, the view makes the wrong predictions about the following Anderson-style subjunctive conditionals (cf. Mackay 2015):

(13) If Jones had taken arsenic, things would not be quite as they actually are.

(14) If Jones had taken arsenic, everything would be exactly as it actually is.

Intuitively, neither of these conditionals is trivially true or false. (13) is true iff Jones did not take arsenic, and (14) is true iff Jones did take arsenic. However, Iatridou's theory predicts that (13) is trivially true and (14) trivially false. According to the theory, a subjunctive conditional quantifies over a domain of worlds that excludes the epistemically possible worlds. But the domain of epistemic possibilities must include the actual world, and thus also every world that makes true all and only the same propositions as the actual world

[8] Iatridou anticipates this worry and suggests fixing the view by adding that 'excludes' be modified to 'excludes and precedes' (in her footnote 19). The suggestion is that, in the modal case, the precedence requirement goes inert. While this technically solves the problem, the addition of precedence renders the exclusion part inert in the temporal case. So, the resulting view seems less of a unified theory that explains the multi-dimensionality of past tense and more like the stipulative "accidental homophony" view it aims to outperform.

(the assumption of factivity for informational modal bases; see also von Fintel & Gillies 2010). Therefore, subjunctive conditionals must quantify over worlds that differ in some respect from the actual world, and hence the consequent of (13) will be true throughout those worlds, and the consequent of (14) will be false throughout those worlds.[9]

We might try to avoid this problem by weakening the exclusion meaning of past tense as follows (cf. von Fintel 1997, which follows Stalnaker 1975):

$T(x)$ is a superset of $C(x)$.

In the modal case, this has the result that subjunctive conditionals quantify over a class of worlds that may reach outside the domain of epistemic possible worlds. This view handles the Mackay problem above, since it allows that in some cases the domain of a subjunctive conditional contains the actual world, which would render (13) false, and in some cases the domain will contain only worlds exactly like the actual world, which renders (14) true. Thus, the view avoids predicting that (13) is trivially true and (14) trivially false.

However, the problem with this fix is that this weakened meaning of past tense is not plausible in the temporal case. Insofar as past-tensed clauses must be throughout the relevant interval quantified over, if this interval had to include the present time, this would ensure that past-tensed clauses entailed that the event described to have taken place in the past was still taking place in the present. In other words:

(15) John was happy.

would entail

(16) John is now happy.

But it seems there is no such entailment. Again, the problem here arises because merely extending the time interval to the past, but

[9] Mackay 2015 shows that the same problem arises for the more complex modal past view of Schulz 2014.

demanding the present time also be within that interval, ensures that we cannot truly describe things as fully past using past tense. But that seems wrong.

This discussion does not exhaust every possible modal past view. For all that's been said, there may be a plausible modal past view that avoids all these problems and generates correct predictions about a range of conditionals. However, by exploring some of the challenges facing the most well-known modal past view, I hope to have raised concerns that motivate us to look at its competitors, temporal past theories. In the next chapter, I will develop a temporal past view that accounts for the semantic/pragmatic differences in indicative and subjunctive conditionals.

8

Temporal Past

8.1 Deriving the Semantic Properties

The key claim of the temporal past view is that the past tense morphology we find on subjunctive conditionals scopes over the conditional and shifts its evaluation time to the past. I will assume for now that this is plausible and can be implemented compositionally given a plausible theory of the morphosemantics of tense (I sketch a possible implementation in the Appendix to this chapter). What is crucial for our purposes is the following claim:

> **Temporal Past**
>
> The time of a subjunctive conditional (at which its covert modal's domain is determined) is earlier than the time at which it is uttered. The time of an indicative conditional is the time at which it is uttered.

Temporal Past is the core commitment of a temporal past theory: it states that the morphological difference between indicative and subjunctive conditionals affects the time at which they are evaluated. I will in addition assume that *will*/*would* are temporal auxiliaries that shift the evaluation time to the future (cf. Comrie 1985, Abusch 1998, Condoravdi 2002, Ogihara 2007). Together, these assumptions generate two crucial predictions about the relationship between conditional time and consequent time in the conditionals (1) and (2):

(1) If John is not in Athens, he is in Barcelona.

(2) If John had not been in Athens, he would have been in Barcelona.

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0009

The first is that the conditional time, consequent time, and utterance time are the same in (1). The second is that conditional time is before utterance time, and consequent time is after conditional time (and perhaps identical to utterance time) in (2).

Clearly, these temporal differences are not enough to account for the semantic and pragmatic differences between indicatives and subjunctives. There is no immediate connection between conditional time being before consequent time and a conditional receiving a metaphysical interpretation, or implying that its antecedent is counterfactual. And neither is there any connection between conditional time being no later than consequent time and the conditional receiving an epistemic interpretation. To remedy this issue, we need a constraint on the assignment of domains to conditionals in context that interacts with the time of the conditional to yield a difference in modal flavor.

Recall that, in the previous chapters, I relativized the extension of a conditional to a context c, modal base m, and a sequence σ. A modal base, recall, is a function from worlds to sets of worlds (the conditional's domain), which was restricted by its antecedent and then fed to σ (which selected a unique closest world to w from that set): $\sigma(m_{\sigma}^{\mathbf{A}})$. I now need to complicate this picture in three ways. First, conditional domains will be determined by two distinct parameters: a modal base m and an ordering source g. Second, both modal bases and ordering sources will be relativized to a world and a time. Finally, both modal bases and ordering sources map a world and time to a set of propositions (rather than a set of worlds). Thus, we have:

- m is a modal base; $m_{w,t}$ is a set of propositions (the value of the modal base at w, t).
- g is an ordering source; $g_{w,t}$ is a set of propositions (the value of the ordering source at w, t).
- D is a domain function that maps a proposition **A** (the conditional's antecedent) and m, g, w, t to a set of worlds: $D^{\mathbf{A}}_{w,t,m,g}$.

Readers familiar with Kratzer 1981, 1991, 2012 will recognize the formal framework I am borrowing from, with D being the **best** function, that delivers the set of worlds compatible with the modal base that are ranked highest by the ordering source (cf. von Fintel 2012a). For now, I will not define $D^{\mathbf{A}}_{w,t,m,g}$ except to say that it must be a subset of the set of worlds compatible with its modal base at its evaluation world and time: $D^{\mathbf{A}}_{w,t,m,g} \subseteq \bigcap m^{\mathbf{A}}_{w,t}$. We will return to the formal definition of $D^{\mathbf{A}}_{w,t,m,g}$ and the contribution of g in Chapter 9. In this chapter, we will focus solely on the role of the modal base.

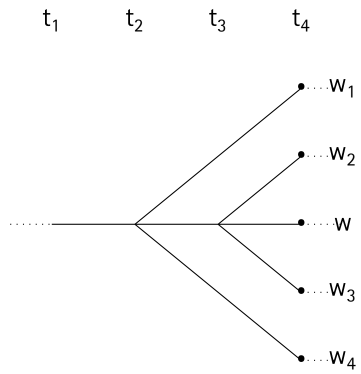
Next, I will make a crucial assumption about the space of possible modal bases. The assumption is that modal bases come in two kinds: those that are **historically constrained** and those that are **informationally constrained**. An informationally constrained modal base i maps a world and time to a set of propositions that comprise the content of some body of information at that world and time (such as what an agent or group of agents know). The informational possibilities at w, t are just those compatible with the output of the relevant informational modal base at w, t: $\bigcap i_{w,t}$. As before, I will assume that informational modal bases are factive, and thus that:

Informational Factivity
For any $w, t : \mathbf{P} \in i_{w,t}$ only if $\mathbf{P}$ is true at w.

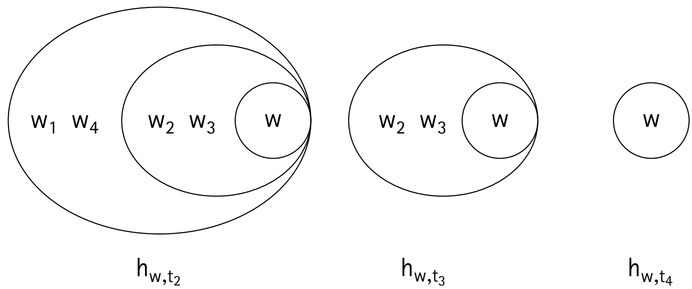
A historical modal base h maps a world and time to the set of propositions that are true at w and about times no later than t. It follows from this that:

Historical Modality
For any w, t and proposition $\mathbf{P}$: $\mathbf{P} \in h_{w,t}$ iff $\mathbf{P}$ is true at w and is about an interval that does not extend beyond t.

The historical possibilities at w, t are $\bigcap h_{w,t}$: the set of worlds that are intrinsically just like w up until t while possibly differing thereafter. We can think of historical possibilities as arrayed in a treelike structure as follows:



As the figure shows, historical possibilities are nested and decrease monotonically over time:



Since historical possibilities decrease monotonically over time, the past (relative to modal time) is settled—whatever is historically necessary now is historically necessary for all past times—whereas the future (relative to that time) may be open—even if it will be later historically necessary that p, it need not now be historically necessary that p.

Notice, then, that there is an important similarity and two important differences between historical and informational possibilities. Both kinds of modal bases are factive; thus, if my conjecture is right, all modal bases are factive. However, they differ in that there is only one historical modal base h, whereas there are many informational modal bases $i_1, \ldots, i_n$, each corresponding to different sources of information. And secondly, while the past (relative to modal time) is historically necessary, this need not be the case for informational possibility—that is, if the relevant agent is ignorant about the past, then the past will not be informationally necessary relative to that agent's information.

So, the first part of my view is the assumption that there are only two kinds of modal bases, informational and historical. I want to be clear that this view is compatible with there being more than two flavors of modality. Indeed, we want to allow that there are also deontic, bouletic, and teleological modals. I handle these by appealing to variation in the ordering source used to determine the modal/conditional's domain (exactly as Kratzer does). Remember, and this is crucial:

Critical point about conditional domains!

The domain of a conditional at t is **not** the set of worlds compatible with its modal base at t. Rather, the domain of a conditional at t is the subset of worlds compatible with its modal base at t that are ranked highest according to its ordering source.

Again, remember, I am setting aside the contribution of ordering sources for the time being (we'll come back to the issue in Chapter 9). But keep this in mind when evaluating the proposal as sketched here.

In particular, the metaphysical flavor of (many) subjunctive conditionals is the result of having a domain determined by a historical modal base h plus a metaphysical ordering source, while the epistemic flavor of (many) indicative conditionals is the result of having a domain determined by an informational modal base i (that not all subjunctives are metaphysical and not all indicatives are epistemic is an issue we will return to in §8.2).

We can now generalize our semantics from the previous chapters to integrate our new relativizations as follows (assume as before that $D_\sigma = D_{\sigma_w}$):

Uniform Semantics

$[\![A \to B]\!]^{m,g,t,\mathbb{Z},\sigma} = 1$ iff $\sigma(D^{\mathbf{A}}_{\sigma,t,m,g} \cap \sigma[\mathbb{Z}]) \in \mathbf{B}$

Roughly, what this says is that the conditional $A \rightarrow B$ is true at $m, g, t, \mathbb{Z}, \sigma$ iff the σ-selected world in its domain (restricted by $\mathbb{Z}$) is a **B**-world. In what follows, I will ignore the role of the partition and the ordering source, and I will assume that subjunctive conditionals involve a past tense scoping above the conditional, and the modal auxiliary *would* heading their consequents (this compositional structure is discussed further in the Appendix to this chapter):

$$past(A \rightarrow would\ B)$$

For now, I will set aside the complication stemming from whether past tense and *will*/*would* involve existential quantification over (past/future) times or select some specific time by putting constraints on which times can be assigned by an assignment function to these free time variables (cf. Partee 1973). I assume for now there is a unique past time at which the subjunctive's domain is determined—call this its conditional time. I will adopt the following terminology as a shorthand:

- t_{c_u} is the time of the context of utterance.
- t_D is the time of the conditional; this is the time relative to which its **domain** is determined.
- t_A is the time the conditional's **antecedent** A is about.
- t_B is the time the conditional's **consequent** B is about.

As we will see momentarily, the crucial relationship is not between t_{c_u} and t_D but between t_D and t_B, the time of the conditional and the time of its consequent.

For now, focus on present indicatives and subjunctives—minimal pairs like:

(1) If John isn't in Athens, he is in Barcelona.

(2) If John hadn't been in Athens, he would have been in Barcelona.

We aim to predict the semantic difference between these in terms of their morphological differences. In particular, we aim to predict that the indicative can only be assigned an informational modal base, whereas the subjunctive can be assigned a historical modal base, and we aim to generate this prediction based on the relationship between their conditional times and consequent times.

To do this, I will again appeal to the Avoid Trivialization principle governing the assignment of context-dependent parameters in context that I motivated in Chapter 5:

Avoid Trivialization

Interpret a complex sentence so that it is not epistemically equivalent with any of its sentential constituents.

Recall that this principle is independently motivated on the grounds that it predicts the infelicity of Hurford disjunctions (Hurford 1974, Katzir & Singh 2013):

(3) a. #John is from Paris, or France.
b. #John is not married, or is a bachelor.

as well as correctly predicts that there's no interpretation of *could A* on which it is epistmically equivalent to A. As we saw in Chapter 6, Avoid Trivialization also rules out interpretations of conditionals $A \rightarrow B$ on which they are epistemically equivalent to their consequents B. As we will show shortly, it is this prediction that will allow us to predict that indicative conditionals like (1) are invariably epistemic, while subjunctive conditionals like (2) may be metaphysical.

Let's start with a statement of the predictions made by Avoid Trivialization, given our assumptions about the structures of informational and historical modal bases and domain functions.

Predictions

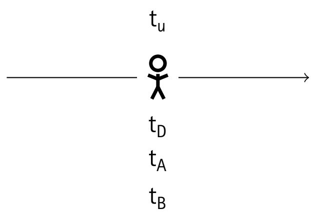
For any conditional $A \rightarrow B$:

(α) If $t_B \leq t_D$, then the conditional must be assigned an informational modal base.

(β) If $t_B > t_D$, then the conditional may be assigned a historical modal base.

These predictions (which we will derive in a moment) align with the observation that the indicative (1) is epistemic and the subjunctive (2) is metaphysical. In (1), the time of the conditional is utterance time, since it has no past tense shifting the evaluation of its conditional time to the past, hence $t_u = t_D$. But the antecedent and consequent of (1) are about the present (where John now is), and hence $t_A = t_B = t_D$:

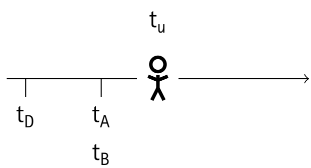
(1) If Smith isn't in Athens, he is in Barcelona.



Thus, we predict that (1) can only receive an epistemic interpretation.

With (2), the time of the conditional is prior to utterance time, $t_D < t_u$ (that is the contribution of the tense morphology, given our assumption Temporal Past). Then, supposing that *would* shifts the time of the consequent to the future of that past time (in the most natural interpretation, back to utterance time), we have $t_B > t_D$:

(2) If Smith hadn't been in Athens, he would have been in Barcelona.



Thus, we predict that (2) may receive either a metaphysical or an epistemic interpretation. Note that we do not yet predict that (2) is most naturally read as metaphysical—this is an issue we will come back to in the next section.[1]

[1] My work here is inspired by a similar move used by Condoravdi 2002 to explain the different readings of modals interacting with past tense. In Khoo 2015, I derived the same predictions by appealing to a diversity condition on modal prejacents and modal bases (which itself was motivated by a principle similar to Avoid Trivialization).

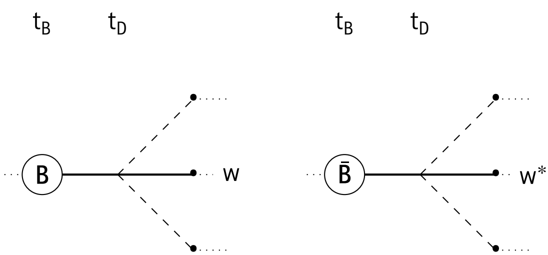
With the predictions stated, I will now show why our theory generates them. Start with (α):

(α) If $t_B \leq t_D$, then the conditional must be assigned an informational modal base.

Let $A \rightarrow B$ be a conditional whose consequent time is no later than conditional time:

$$t_B \leq t_D$$

and suppose for reductio that context assigns this conditional a historical modal base h. It follows that $A \rightarrow B$ is equivalent to B (assuming both are defined), contradicting Avoid Trivialization.[2] To see this, try to imagine what it would be like for B to be true and $A \rightarrow B$ false or vice versa. The situation would look like this:



On the left, we have **B** is true at w and on the right $\bar{\mathbf{B}}$ is true at w^*. Since B occurs before the time of the conditional t_D, **B** is historically necessary at w at that time, while $\bar{\mathbf{B}}$ is historically necessary at w^* at that time. Now, if $A \rightarrow B$ is assigned a historical modal base, this means that it is true iff all of the **A**-worlds in its domain (which, remember, is a subset of the historically accessible worlds) are **B**-worlds. Thus, since all of the historically accessible worlds from w at

[2] Here, I simplify the reasoning by just looking at factual truth of a conditional content at a world, rather than a sequence of worlds. Note that a conditional is factually true at a world iff it is true at every sequence starting with that world; it is factually false at a world iff it is false at every sequence starting with that world; and it is non-factual otherwise.

t_D are **B**-worlds, $A \rightarrow B$ is factually true at w. And since none of the historically accessible worlds from w^* at t_D are **B**-worlds, $A \rightarrow B$ is factually false at w. Thus, we have proved that, for any conditional whose consequent time is no later than conditional time, if it is assigned a historical modal base, it will be equivalent to its consequent, thus violating Avoid Trivialization. But since informational modal bases are not historically constrained in this way, there will be informational modal bases such that assigning them to $A \rightarrow B$ does not make it equivalent to its consequent. And this establishes (α).

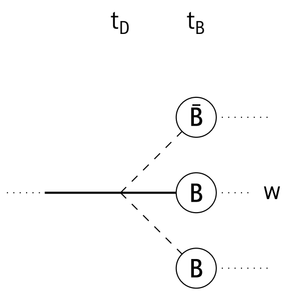
Turn next to (β):

(β) If $t_B > t_D$, then the conditional may be assigned a historical modal base.

Let $A \rightarrow B$ be such that its consequent time comes after its conditional time:

$$t_B > t_D$$

Suppose that context assigns this conditional the historical modal base h. This need not lead to a violation of Avoid Trivialization. This is easy to see in the following graphic:



Here, since t_D comes before t_B, the historically possible worlds may differ in the status of **B** even though **B** is in fact true.

To summarize, we have seen that, given Historical Modality and Avoid Trivialization, we predict (α) and (β). Furthermore, given Temporal Past, and our assumption that *will*/*would* shift evaluation time, we immediately predict that (1) must be assigned an informational modal base and (2) may be assigned a historical modal base (or an informational modal base):

(1) If John is not in Athens, he is in Barcelona.

(2) If John had not been in Athens, he would have been in Barcelona.

Thus, we predict that past and present indicatives must be epistemic, whereas past, present, and future subjunctives may be either metaphysical or epistemic. And, furthermore, from (β) we also predict that future indicatives like (4) may be either metaphysical or epistemic:

(4) If John does not go to Athens, he will go to Barcelona.

Summarizing our predictions:

	Indicative	Subjunctive
Past	Epistemic	Epistemic/Metaphysical
Present	Epistemic	Epistemic/Metaphysical
Future	Epistemic/Metaphysical	Epistemic/Metaphysical

In the next two sections, I argue that these predictions are correct: the indicative/subjunctive grammatical distinction crosscuts the epistemic/metaphysical semantic distinction.

8.2 Non-Standard Predictions

8.2.1 Epistemic Subjunctives

Interestingly, epistemic readings of subjunctive conditionals have already been observed, although they have mostly been dismissed as oddities.[3] Here are two examples from Edgington 2007 (the first she borrows from Grice 1989, and the second from van Fraassen 1981):

[3] See Veltman 2005, Schulz 2007, Kratzer 2012.

Treasure Hunt. "There is a treasure hunt. The organizer tells me:

(5) I'll give you a hint: it's either in the attic or in the garden.

Trusting the speaker, I think

(6) If it's not in the attic, it's in the garden.

We are competing in pairs: I go to the attic and tip off my partner to search the garden. I discover the treasure. 'Why did you tell me to go to the garden?' she asks. [I reply:]

(7) Because if it hadn't been in the attic, it would have been in the garden: that's what I was told. (Or more pedantically: 'that's what I inferred from what I was told')

That doesn't sound wrong in the context."

Skeptical Child. "The conjuror holds up a penny and claims he got it from the boy's pocket. 'That didn't come from my pocket,' says the boy. 'All the coins in my pocket are silver. If that had come from my pocket, it would have been a silver coin.'" Edgington 2007: 212–213

Both of these subjunctives seem true as uttered in these contexts:

(8) If the treasure hadn't been in the attic, it would have been in the garden.

(9) If that coin had come from my pocket, it would have been silver.

Yet, they seem true merely because the speaker's epistemic state at some past time entailed $A \supset B$. This is an instance of the or-to-if inference, which does not hold in general for subjunctives:

(10) a. Either Oswald shot Kennedy, or someone else did.
b. #So, if Oswald had not shot Kennedy, someone else would have.

On its metaphysical interpretation, a subjunctive's (factual) truth requires some feature of the world to ground it (perhaps the presence of a backup shooter in the case of (10-b)). Notice as well that both (8) and (9) have standard metaphysical readings as well. On its metaphysical reading, the truth of (8) requires some fact about the treasure hider—perhaps that they planned to use the garden as a backup hiding place for the treasure or maybe that the treasure was initially hidden in the garden and changed at the last minute to be hidden in the attic. Similarly, on its metaphysical reading, the truth of (9) requires some unusual law connecting the nature of the coin with its being drawn from my pocket. Important for our purposes is the fact that, on the interpretations most natural in their respective contexts, neither (8) nor (9) require any such thing for their truth. Rather, all they seem to require for their truth is that the relevant past epistemic state (in this case, the speaker's) entailed the corresponding material conditional. This is strong evidence that these conditionals are receiving a past epistemic interpretation in these examples, and thus evidence that there can be epistemic interpretations of subjunctive conditionals.

But now we come to an obvious question. Since these epistemic interpretations took quite a bit of contextual priming to bring out, it seems the default interpretation of a subjunctive conditional is metaphysical. Why is that? My answer is that learning about relations between past epistemic possibilities that are no longer open is only relevant to questions about past states of knowledge. However, since relations between past metaphysical possibilities are grounded in other facts (like intentions, plans, laws, etc.), learning about them is relevant to such factual questions, as well as our plans for future courses of action. Thus, from the fact that information about past epistemic possibilities is relevant only to a much narrower set of questions than information about past metaphysical possibilities, we expect the default interpretation of a subjunctive conditional to be metaphysical. In other words, if you are not sure what question is at issue in the context, but you want an utterance of a subjunctive conditional to be relevant to that question, you are more likely to get a relevant interpretation if you interpret the subjunctive as metaphysical.

However, we also expect this default interpretation to be overridden in contexts in which information about past epistemic states is clearly relevant. Such contexts include ones in which the speaker is explaining or evaluating her past behavior: there, her past epistemic state is clearly relevant, since those were the possibilities that she used to guide her actions. And, we do seem to find epistemic subjunctives in exactly those contexts. In **Treasure Hunt**, the speaker is explaining why she told the hearer to check the garden, and in **Skeptical Child**, the speaker is explaining why he does not believe the conjurer's claim. Further confirmation of this prediction is the observation (due to von Fintel & Gillies 2008) that *might have* can also receive a past epistemic interpretation as *might have* does in B's second response to A:[4]

(11) A: Where are the keys?
B: They might be in the car.
A: (after checking) They are not, why did you say that?
B: I didn't say they were in the car, just that they might be—and they might have been!

I conclude, then, that subjunctive conditionals can receive both metaphysical (by default) and epistemic interpretations. This confirms an important prediction of my theory.

8.2.2 Metaphysical Indicatives

The second unexpected prediction of my theory is that future-oriented indicatives (those whose consequents are about a time to the future of utterance time) may have both epistemic interpretations and metaphysical interpretations. Nonetheless, I think it is correct.

[4] This is in addition to its present epistemic (but past-looking) and past metaphysical interpretations (cf. Condoravdi 2002):

(i) The keys might have been in the car.
a. It's now possible that they were in the car. — Present epistemic
b. It was then possible that the keys ended up in the car. — Past metaphysical

Suppose that I know a particular mushroom is lethally poisonous. On that basis, I tell you:

(12) If you eat that mushroom, you will die.

You trust me and don't eat the mushroom. When you throw it away, I am relieved and say:

(13) If you had eaten that mushroom, you would have died.

In this case, it seems that what makes (12) true also makes (13) true—the fact that the mushroom is poisonous. Intuitively, the two stand together. They also fall together. To see why, suppose I think the mushroom is poisonous and I tell you (12), and on that basis you throw it away. However, now suppose that you find out later I was wrong about the mushroom being poisonous. It seems that you might in that case respond to me by saying:

(14) If I had eaten that mushroom, I wouldn't have died.

Intuitively, your statement here contradicts mine: thus, the future indicative and its subjunctive counterpart (uttered after the antecedent time) seem to stand or fall together. I claim that in such a scenario, (12) receives a metaphysical interpretation, since its truth intuitively stands or falls with the same fact that the metaphysical interpretation of its corresponding subjunctive stands or falls with (see Edgington 1995, Bennett 2003 for further discussion of cases like these).

Now, consider a third version of this scenario. In this case, I know the mushroom before you is one of the *fungi terriblis*—these are not poisonous, but I know them to have a curious property: all and only the people who eat them in fact have a pre-existing medical condition, sporesis, that is triggered by the mushroom, causing them to die shortly thereafter. So, learning that you have eaten a *fungi terriblis* is really bad news, although it alone does not kill you. I have no idea whether you have sporesis, but on the basis of what I do know, I say (12), and, it seems, my assertion is correct (both in that I am warranted in asserting it, and that I say something true). You discard the mushroom. Later, you are screened by your doctor and you learn that you do not have sporesis. On that basis, you say, (14). Intuitively, in this case, the truth of (14) does not contradict what I said when I uttered (12) earlier. After all, the truth of (12) (as uttered by me earlier) did not depend on whether you in fact had sporesis—I might have thought it very unlikely that you did (and thus very unlikely that you would have eaten the mushroom) and still correctly asserted (12). That is unexpected if (12) (in that context) were true iff (14) (uttered by you later) were true, since the latter does seem to hold just in case you have sporesis.

In this third version of the scenario, it seems that (12) does not stand or fall with (14). I claim that in this version, it has an epistemic interpretation, whose truth does not depend on the fact that grounds the truth of its corresponding metaphysical subjunctive (14).

That is the prima facie case in favor of the claim that future indicatives have both metaphysical and epistemic interpretations. For other pairs of epistemic future indicatives that do not stand or fall with their metaphysical subjunctive cousins, see Schulz 2017: 43–44.

8.2.3 How Are the Two Readings Related?

We have seen the case for divergent interpretations of future indicatives and past subjunctives. I now want to focus on the relationship between the epistemic and metaphysical readings themselves. Minimal pairs like (15)/(16) suggest that there is a truth-conditional difference between epistemic and metaphysical readings of conditionals.

(15) If Oswald didn't shoot Kennedy, someone else did.

(16) If Oswald hadn't shot Kennedy, someone else would have.

Intuitively, (15) seems plausibly true, or indeed something we might reasonably be sure of on the grounds that we know someone shot Kennedy. By contrast, (16) seems plausibly false or at least something we might reasonably think is unlikely, on the grounds that we think it is likely there was no backup shooter.

To see a reverse case where the subjunctive is more probable than the indicative, consider the following:

> **Jewels**. The jewels are missing. There's evidence that it was Smith who stole them, but the only other suspect, Jones, has a strong alibi. Given that Smith did not do it, it's more likely that the jewels were misplaced than that anyone stole them. On that basis, it is unlikely that:
>
> (17) If Smith did not steal the jewels, someone else did.
>
> However, it's also known that Smith and Jones work in a team, and that they were targeting these jewels. Usually, in that case, one of them serves as a backup in the event that the other fails (or is caught). On that basis it is likely that:
>
> (18) If Smith had not stolen the jewels, someone else (namely Jones) would have.

Thus, it seems epistemic past indicatives are truth conditionally orthogonal to their metaphysical subjunctive counterparts (uttered at the same time). Note that our theory predicts this, given a notion of classical entailment as truth preservation of refined content across sequences:

Classical Entailment
$\mathtt{A}_1, \ldots \mathtt{A}_n \vDash \mathtt{C}$ iff for all $\sigma \in \mathtt{A}_1, \ldots \cap \ldots, \mathtt{A}_n : \sigma \in \mathtt{C}$

We predict this because the domains of indicatives and subjunctives may be orthogonal, so choosing an **A**-world that is a **B**-world from one domain might not involve similarly choosing an **A**-world that is a **B**-world from the other domain.

However, there is another sense in which epistemic indicatives and metaphysical subjunctives pattern more closely together. Contrast the following Newcomb scenarios (Nozick 1969). Start with the initial setup. Smith is presented with a choice between taking the contents of an opaque box, which contains either \$0 or \$1,000,000, and taking the contents of the opaque box and a clear box that contains \$1,000. Whether there is \$0 or \$1,000,000 in the opaque box depends on what a reliable computer model of his psychology predicted he would take.

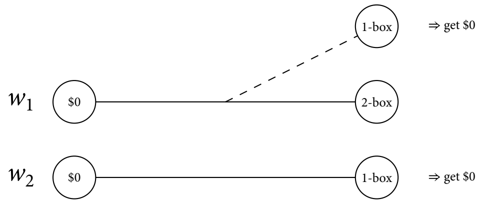
In the first version of the case, you are not sure what Smith chose (one box or two) but you believe the computer is reliable and you also think it is likely he chose two boxes (him being rational and all). The following sentences seem correct:

(19) Probably, if Smith had taken one box, he would have got \$0.

(20) Probably, if Smith took one box, he got \$1,000,000.

Smith probably took two boxes, in which case the computer probably predicted he would and put \$0 in the opaque box. But then if he had instead taken one box, he would have probably got \$0 since that box would have been empty. Of course, once you conditionalize on the fact that Smith took one box, you update your credences and, continuing to think the computer was reliable, think the computer probably put \$1,000,000 in the opaque box, in which case Smith probably won \$1,000,000.

In the second version of the case, suppose now that you know the opaque box is empty, but not whether the predictor was right or what Smith chose. Thus, you know that either Smith chose one box and the predictor was wrong (since there is nothing in that box), or he chose two boxes and the predictor got it right (but since the other box was empty, had he chosen the other box, he would have got \$0). Your information state is made up of worlds like w_1, w_2:



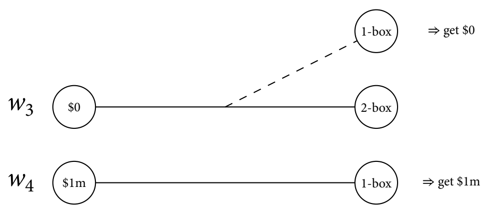
On that basis, it seems that you know both (21), and on that basis (22):

(21) If Smith had taken one box, he would have got \$0.

(22) If Smith took one box, he got \$0.

In this context, it seems that you believe the subjunctive (21) only if you believe the corresponding indicative (22). Supposing you do not believe the indicative—that is, you allow for the possibility that Smith took one box and got \$0, then it seems you cannot (fully) believe that if he had taken one box, he would have got \$0.

Does the reverse hold? Can you believe the indicative without believing the subjunctive? Consider the following, third, version of the case: now you know the predictor was right, but not whether the opaque box was empty or what Smith chose. Then, your information state is made up of worlds in which the opaque box is empty and Smith chose both boxes and worlds in which the opaque box has \$1,000,000 and Smith chose one box:



On that basis, it seems you know (23), since every epistemically possible world in which Smith chose one box is a world in which he got \$1,000,000. But it would be wrong to infer (24) since not every epistemically possible world is one in which the closest metaphysically possible world in which Smith took one box is one in which he got \$1,000,000 (w_3 above is a counterexample, since it is compatible with your information that Smith took both boxes, and there was no money in the opaque box).

(23) If Smith took one box, he got \$1,000,000.

(24) If Smith had taken one box, he would have gotten $1,000,000.

Generalizing:

> **Observation 1**
> If you believe a metaphysical subjunctive and leave open the possibility of its antecedent, then you believe its epistemic indicative counterpart.
> **Observation 2**
> It is possible to believe an epistemic indicative and not believe its metaphysical subjunctive counterpart.

We predict both observations. Take Observation 1. If an agent leaves open **A** and believes the subjunctive $A \Rightarrow B$, then she cannot allow any $\mathbf{A\bar{B}}$-worlds in her belief state (since by Modus Ponens $A \Rightarrow B$ is false at such worlds). But that is sufficient for her to believe the indicative $A \rightarrow B$, as long as the indicative's modal base is bounded by her beliefs.[5]

[5] Note that the restriction to leaving open the possibility of the antecedent is crucial in Observation 1. Indicative conditionals exhibit what I will call "indicative context shift" (cf. Weatherson 2001, Santorio 2012, Mackay 2013)—this means that the world of the context of indexicals appearing in the antecedent or consequent of an indicative conditional is shifted, pointwise, to each world in the conditional's domain. As a result, conditionals like the following either suffer from presupposition failure or are necessarily false:

(i) a. #If John is taller than he (actually) is, he is able to reach that lightbulb.
b. #If John reached the lightbulb, he is taller than he (actually) is.

By contrast, subjunctives do not exhibit "indicative context shift," which means that indexicals appearing in their antecedent/consequent remain evaluated in the actual context of utterance. This allows the subjunctive counterparts of (i-a)/(i-b) to avoid presupposition failure and be non-trivially true/false:

(ii) a. If John were taller than he (actually) is, he would have been able to reach that lightbulb.
b. If John had reached that lightbulb, he would have been taller than he (actually) is.

(ii-b) could be accepted in a context in which (i-b) is defined but false throughout. My account does not explain this difference between indicative and subjunctive conditionals. However, it should be noted that the problem is not specific to conditionals; other past-tensed operators allow for similar contrasts (cf. Russell 1905):

(iii) a. #I think you are taller than you (actually) are.
b. I thought you were taller than you (actually) are.

Take Observation 2 and suppose the agent's belief state comprises **AB**-worlds and $\bar{\mathbf{A}}$-worlds. Then, the agent will believe the indicative conditional $A \rightarrow B$ (given that its modal base is bounded by her beliefs). But this is compatible with some of the $\bar{\mathbf{A}}$-worlds in her belief state being such that all of metaphysically accessible **A**-worlds to them are $\bar{\mathbf{B}}$-worlds, which makes the subjunctive $A \Rightarrow B$ false at them.

Summary: *Indicative and subjunctive*

Epistemic indicatives are truth conditionally orthogonal from their metaphysical subjunctive counterparts. One can believe an epistemic indicative without believing its metaphysical subjunctive counterpart. But one cannot believe a metaphysical subjunctive without believing its epistemic indicative counterpart.

8.3 Objections and Replies

In this section, I consider and reply to several challenges facing the temporal past view, and my particular strategy for deriving the semantic differences between indicative and subjunctive conditionals.

8.3.1 Backwards Subjunctives

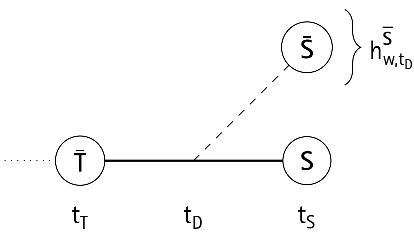
A crucial prediction of my view is that metaphysical conditionals and modals must be such that their consequents/prejacents are about times later than conditional/modal time. This is to ensure that the resulting conditional/modal meets the Avoid Trivialization constraint, in accordance with the fact that propositions about the past are historically necessary. Focusing on the relationship between conditional time and consequent time marks an important difference between my historical modality theory of (metaphysical) subjunctives and the standard way of formulating such theories. Standard formulations of branching time semantics for subjunctive conditionals are about the relationship between antecedent and consequent time (see the discussions in Lewis 1979c, Bennett 2003):

> **Standard Branching Theory**
> $A \Rightarrow B$ is true at w iff B is true in all of the A-worlds that are exactly like w up until just before the time of A (and which possibly diverge thereafter).

This theory immediately runs into problems with backwards subjunctives: these are subjunctives whose consequent is about a time before its antecedent, such as,

(25) If Oswald hadn't shot Kennedy, he would have been talked down by the CIA.

Lewis 1979c raises this example as a problem for the Standard Branching Theory. That theory is forced to predict that backwards conditionals like (25) are false simply because their consequents are false. This is because every world exactly like ours up until just before the time at which Oswald is hypothesized to not have shot Kennedy is one in which he was not talked down by the CIA:



Every world branching at t_D (just before $t_{\bar{S}}$) in which Oswald did not shoot Kennedy is still one in which he was not talked down by the CIA (since this is settled before t_D)—and this is enough for the Standard Branching Theory to deliver the verdict that (25) is false.[6]

[6] A possible reply on behalf of the theory is to propose that there are no such worlds. If that is correct, then the theory does not predict that (25) is false, but rather that it is trivially true. This is not obviously a better result, but we can adjust the example to make

Lewis's problem is a real one for the Standard Branching Theory, but let me emphasize: **my theory is not the Standard Branching Theory**! Rather, on my theory, we hold fixed the facts up until some time prior to the time of the conditional, which is not necessarily some time immediately prior to the antecedent:

> **Metaphysical Subjunctives**
> A (metaphysical) subjunctive conditional $\mathtt{A} \rightarrow^{\mathtt{t}}_{\mathtt{h}} \mathtt{B}$ is factually true at *w* iff *B* is true at all the *A*-worlds that are exactly like *w* up until just before the time of the conditional *t* (and which possibly diverge thereafter).

This allows my theory in some cases to select a conditional time that is much earlier than the antecedent—this is what I think happens with backwards subjunctives like (25). But do I predict this? Yes! Given Historical Modality and Avoid Trivialization, we require that the conditional time for a metaphysical subjunctive must be prior to *consequent* time. For backwards subjunctives, since their consequent time precedes their antecedent time (see the diagram above), this means the conditional time will come before both consequent and antecedent time. This is what allows the theory to "undo" the truth of the consequent in evaluating a backwards conditional.

Note, however, that I have not yet said anything about the relationship between antecedent time and conditional time. In the next chapter, I will argue that there is default interpretive pressure to find a conditional time that is before antecedent time as well.

our point in a slightly different way. Take a subjunctive with a false antecedent and true consequent, where the consequent time precedes antecedent time:

(i) If Oswald hadn't shot Kennedy, he (still) would have been in the book depository earlier.

I take it that there are interpretations of this conditional that are not necessarily true. However, the SBT cannot predict them. Holding fixed things until just before the hypothesized non-shooting at *t* is to hold fixed that Oswald is in the book depository just before *t*, in which case the conditional is predicted to be true. But if there are no non-shooting worlds that hold things fixed until just before that time where Oswald is in the book depository just prior, then the conditional is predicted to be trivially true.

8.3.2 Temporally Unusual Antecedents/Consequents

Not all subjunctive conditionals are like (2)—about events or states that happen at various points in the (possible) history (or future) of the world. Subjunctive conditionals with antecedents and consequents that are about no particular times raise a potential problem for my view. The worry is that, if the consequent of a subjunctive conditional is not about any particular time, then, supposing the consequent is false, it will be false at every historically possible world, in which case my theory predicts that the conditional cannot (by Avoid Trivialization) receive a metaphysical interpretation. Consider:

(26) If time had been infinite, then time would have had no beginning.

Let '**N**' denote the proposition expressed by (26)'s consequent (that time had no beginning). It may seem that this proposition cannot be historically contingent, since, to find a time at which it is would require finding a historical branch time before all time.

However, this thought is actually incorrect. Recall the definition of Historical Modality:

> **Historical Modality**
> For any w, t: $\mathbf{P} \in h_{w,t}$ iff **P** is true at w and about an interval that does not extend beyond t.

The propositions that make up the historical modal base at w, t are those about times not extending beyond t. But, it seems that either **N** is not about any interval of time (since it is atemporal) or it is about the entire history of time (since it is true at every time). In the first case, since it is not about any interval, it is not about any interval extending beyond t. But then **N** may be true and not in $h_{w,t}$, for any t. In the second case, since it is about the maximal interval, for any past time t, it is about an interval that extends beyond t, in which case it may be true and not in $h_{w,t}$. Thus, in either case, there is no problem finding a metaphysical interpretation of (26) compatible with Avoid Trivialization.

However, a slightly different problem is not so easily avoided on my view. This problem arises for conditionals whose antecedent or consequent is about the first time ($\exists! t : \neg\exists t' < t$):

(27) If the initial state of the universe had been different, then everything would have been different than it is at all times.

Let '$\mathbf{I}^w$' denote the proposition expressed by the antecedent of (27)—that the initial state of the universe was different (than it is at w). This proposition is about the world of evaluation w, and hence false of it—thus, its negation is true of w. Furthermore, since it is entirely about the first time, for any time t, $\mathbf{I}^w$ is about an interval that does not extend beyond t. Therefore, $\bar{\mathbf{I}}^w \in h_{w,t}$, for any t. But this means that (27) will be undefined, since there are no worlds in $\mathbf{I}^w \cap \bigcap h_{w,t}$. Yet, intuitively, (27) may be true—at the very least, it doesn't strike us as a conditional that is undefined.

I think conditionals like (27) do pose a real challenge to my view. But I don't think that the challenge is insurmountable. Here are two natural fixes. The first amends **Historical Modality** to avoid this problem. Let t^*_w be the first moment of time of w, if there is one. Then, we have:

Historical Modality
For any world w, time $t \neq t^*_w$: $\mathbf{P} \in h(w,t)$ iff $\mathbf{P}$ is true at w and about an interval that does not extend beyond t.
For any world w: $h(w, t^*_w) = \varnothing$.

This strategy predicts that the historical possibilities at the first moment of time are all those worlds compatible with any setting of the initial conditions. This allows us to interpret (27) non-trivially with respect to t^*_w, in which case it would be assigned an empty modal base, restricted by $\mathbf{I}^w$. Such a conditional may be true (it will depend on how things go with the laws of nature, which will come in via the ordering source).

A second possibility is to suppose that (27) is strictly speaking undefined, but that it is possible to interpret it as defined by engaging in a bit of pretense by pretending that time is embedded in a fictional super-time. Within that fiction, there is a conditional super-time at which it is historically contingent what the initial (in the normal sense of time) state of the universe is.

I find both strategies reasonable, and I want to allow that we could, in principle, use either in evaluating odd subjunctive conditionals like (26). A similar challenge comes from subjunctives with impossible antecedents—I will discuss these in Chapter 9.

8.3.3 The Problem of Hindsight

Suppose that coin flips are random events—nothing in the laws and setup of a flip necessitates its outcome. I bet that the coin will land heads. You flip the coin and it lands tails. I regret my decision to bet heads, thinking,

(28) If I had bet tails, I would have won.

This seems right, but notice that if we meddle with the past, changing how I bet, and then rerun the toss again, there's no guarantee the coin will land as it actually did. It seems, then, that there are historically possible worlds stemming from the past branch point before my bet at which the coin lands heads, and ones at which the coin lands tails. So, (28) cannot be (determinately) true.

This is the problem of hindsight (see Slote 1978, Edgington 2004). And, given what I've said so far, it is a serious problem for my view. I take up this problem in detail in Chapter 9. But let me just note here the rough outline of how I propose to solve it. What hindsight shows us is that the relevant subset of historically possible worlds must all make true actual facts that are causally independent of the conditional's antecedent. In other words, if **P** is true and caused by **Q**, which is causally independent of **A**, then **P** will also be true at all of the relevant historically accessible **A**-worlds that comprise the domain of the subjunctive conditional. In the case above, the outcome of the toss is causally independent of my bet, which is why it must continue to hold at the worlds in the conditional's domain.

8.3.4 Wishes and Weak Necessity

Some modal past theorists (in particular Iatridou 2000) are motivated not just by conditionals, but by the fact that the same morphology is co-opted in *wish*-talk, which is obligatorily counterfactual, and weak necessity modals:[7]

(29) I wish I had gone to the party (# and I did).

(30) You ought to go to the party.

Iatridou observes the following generalization about counterfactual *wish* across a range of languages: the distinctive morphology appearing on the antecedent of a subjunctive conditional appears in the complement of *wish*, and the distinctive morphology appearing on the consequent of a subjunctive conditional appears on the *wish*-verb. In other words, where English has a dedicated morpheme *wish* to express counterfactual desire, other languages use the same fake past morphology as on the consequents of subjunctive conditionals. Here is an example from Hungarian (see Iatridou 2000 for more examples):

(31)	Ha	János	tud**ná**	a	választ,	Mari	is	tun**ná**
	If	János	know.**NA**	the	answer-acc,	Mari	too	know.**NA**
	a	választ.						
	the	answer-acc.						

(32)	Szeret-**né**-m	ha	magasabb	len-**ne**.
	like-**NA**-1sg	if	taller	be-**NE**.
	I wish she was taller.			

[7] Wishes are obligatorily counterfactual to the subject's information state, *not* the speaker's (cf. Iatridou 2000: 243–244). We can see this with an example where the subject and speaker are distinguished:

(i) Smith wishes he had gone to the party, but he doesn't remember that he was there.

This difference between wishes and subjunctive conditionals (which are counterfactual, if at all, to the speaker's information state) remains an open issue for further investigation.

In English, following the Hungarian as closely as possible would yield:

(33) I would want it that I had gone to the party.

Here, the past tense on *want* is fake—the speaker desires actually and now that she had gone to the party. And, like with fake tense on conditionals, what is desired is obligatorily counterfactual to the subject's belief state—it signals that the subject has ruled out the possibility of the complement obtaining. As a result, *wish*-claims about unsettled future events are odd, as compared with present futurates (which are settled at the present time):

(34) #I wish I would go to Disneyland next year.

(35) I wish I were going to Disneyland next year.

Weak necessity modals like *ought* are distinguished from their strong necessity counterparts (like *must* or *have to*) by the fact that strong necessity entails weak but not vice versa:

(36) a. #You have to go to the party, but it is not the case that you ought to.
b. You ought to go to the party, but you do not have to.

von Fintel & Iatridou 2008 argue that in many languages weak necessity is expressed by fake past morphology on top of a strong necessity modal. For instance, in Hungarian we have:

(37) Péter-nek el kell-**ene** mosogat-ni-a az edény-ek-et.
Peter-DAT PRT must-**NA** was-INF-3sg the dish-PL-ACC.
Peter ought to do the dishes.

These observations raise trouble for the temporal past view, assuming there is a unified explanation for all these instances of fake past tense. The problem is that it is not obvious that shifting the time of the wish or the modal to the past would generate these results. In the case of *wish*, recall above that *wish*-claims are claims about the subject's actual and present desire for some counterfactual event, and similarly *ought*-claims are about the subject's actual and present weak obligations. Thus, a temporal past theory, on which such claims should involve some genuine past tense on top of a desire verb or necessity modal, seems to be in trouble.

I cannot answer this objection in full here. This is, after all, a book on conditionals, and the temporal past just one of many theses I am defending. However, I want to sketch here a strategy someone sympathetic to a temporal past theory might pursue for counterfactual *wish* and weak necessity modals. The basic idea is that desire verbs and modals have two temporal parameters, a temporal orientation and temporal outlook. The temporal orientation of a desire verb is the time at which the desire is said to hold. The temporal outlook of the desire verb is the time at which the domain of possibilities it ranks is determined. If we think of desire verbs as quantifying over the "best"-ranked worlds from some set of alternate possibilities, the features relevant to the ranking (the ordering source in Kratzer's sense) are determined relative to the temporal orientation, whereas the possibilities that are ranked (the modal base in Kratzer's sense) are determined relative to the temporal outlook. Thus, the domain of a conditional $A \to B$ is:

> $D^{\mathbf{A}}_{w,t,t',m,g}$, where t is the conditional's temporal outlook and t' its temporal orientation.

Then, my hypothesis is that real past tense shifts the temporal orientation (ordering source time), whereas fake past tense shifts the temporal outlook (modal base time). This predicts the contrast between:

(38) a. I wish I had gone to the party. — Lexicalized fake past on *want*
b. I once wanted to go the party. — Real past on *want*

We predict that weak necessity modals differ from genuinely past-tensed modals in that the former's temporal outlook is past-shifted (which, assuming they are assigned a historical modal base, allows them to access past possibilities presently foreclosed), while the latter's temporal orientation is past-shifted (allowing the output of its ordering source to differ from the present time):

(39) a. John ought to have washed the dishes.
b. John was once obligated to wash the dishes.

This is just a sketch of a strategy for building a compositional temporal past theory for counterfactual desire verbs and weak necessity modals. I have not provided any reason to think that such a theory could even be plausibly implemented compositionally. But I am hopeful. The distinction between temporal orientation and temporal outlook seems quite intuitive to me. And, finally, it is worth noting that there is currently no serious compositional modal past theory to compete with. This is truly a wide open issue, and I expect important work to be done on it in the coming years.

8.4 Temporal Past: Deriving the Pragmatic Differences

We come finally to the pragmatic difference between indicative and subjunctive conditionals: typically (but not always), past subjunctive conditionals are counterfactual—a speaker uttering one conveys that they think its antecedent is false—whereas indicative conditionals are always open—a speaker uttering one conveys that they think its antecedent is epistemically possible. What accounts for these pragmatic differences?

To start, notice that we already have the resources to predict that past indicative conditionals will always be open. This is because, if uttered in a context whose information entails $\bar{\mathbf{A}}$, a past indicative $A \rightarrow B$ will be undefined at every world in that information state. We expect this semantic presupposition to generate a pragmatic presupposition by way of the following bridge principle (cf. Strawson 1950, von Fintel 2004):

Presupposition Bridge
A speaker assertively uttering A at c will believe that $[\![A]\!]^c$ is defined at every σ compatible with the information of c.

This principle is rationalized by what it is to make an assertion. Recall that to assertively utter a sentence is to aim to update the context so that its information entails that the sentence is determinately true. This involves ruling out every world from the contextual information at which the sentence is determinately false. But this procedure can only be carried out if the sentence's content is defined at each world in the information of the context. Thus, we expect speakers to believe these preconditions are in place when assertively uttering a sentence, which is what Presupposition Bridge states. But then it follows from the fact that a past indicative $A \rightarrow B$ is defined only if the information of the context contains some **A**-worlds where the speaker uttering it will believe that **A** is possible given the information of the context.

Turn next to (past) subjunctives. First, notice that we predict that non-counterfactual subjunctives are possible. This is because a subjunctive conditional may have a true antecedent and still be true. A (metaphysical) subjunctive conditional is evaluated at a past time t, and is (determinately) true just if all of the relevant historically possible worlds at t in which its antecedent is true are worlds at which its consequent is true. Since w is always one of its own historical possibilities, a subjunctive with a true antecedent is (determinately) true only if its consequent is also true. This is a plausible prediction. Recall Anderson 1951's non-counterfactual subjunctive:

(40) If Jones had taken arsenic, he would have shown just exactly those symptoms that he does in fact show.

As Anderson notes, (40) may be put forward as evidence that Jones took arsenic. According to our theory, (40) is (determinately) true just if all of the relevant historically possible arsenic-worlds (from the relevant time) are ones in which Jones shows just the symptoms he shows now. Suppose Jones in fact exhibits symptom X. A background assumption is that it is a consequence of some relevant laws of nature that if Jones took arsenic, he exhibits symptom X (or, alternatively, that taking arsenic is a cause of symptom X). Since the relevant historically possible worlds are ones that match our laws (something we will discuss in the next chapter), it follows from the (determinate) truth of (40) that at all the relevant historically possible worlds where Jones took arsenic, he exhibits symptom X. And since Jones in fact exhibits symptom X, the truth of (40) is confirming evidence that Jones took arsenic. Of course, the truth of (40) is also compatible with Jones not taking arsenic, as long as there are other causes of symptom X. Nonetheless, our theory accounts for why (40) is evidence that Jones took arsenic.

Similarly, our theory predicts that Mackay 2015's conditionals (41)/(42) are not trivially true/false:

(41) If Jones had taken arsenic, things would not be quite as they actually are.

(42) If Jones had taken arsenic, everything would be exactly as it actually is.

Recall that Iatridou's theory made the wrong predictions about these subjunctives because, according to her theory, subjunctives must quantify over worlds that exclude the epistemically possible ones, thus excluding the actual world. However, we have already seen that our theory allows that subjunctives can quantify over the actual world (if their antecedent is true there). And thus, we predict (correctly) that (41) is true iff Jones didn't take arsenic, and that (42) is true iff Jones took arsenic.[8]

8.4.1 Leahy's Antipresupposition Theory

In a series of papers (Leahy 2011, 2016, 2018), Brian Leahy proposes a theory on which the counterfactual signal generated by subjunctive

[8] One half of each of these predictions goes through only if a subjunctive with a true antecedent is true iff its consequent is also actually true. We get this result by the *Minimality* condition of selection functions.

conditionals is an antipresupposition. Here, I will articulate a challenge to Leahy's strategy.

An antipresupposition is an implicature that is generated by the use of a sentence A that has a lexical alternative B, which carries a stronger presupposition A (hence, the presupposition of B, π_B, asymmetrically entails the presupposition of A, π_A). The implicated content is that B's presupposition is false, $\neg\pi_B$. One source of this implicature is the principle Maximize Presupposition (cf. Heim 1991, Sauerland 2008, Schlenker 2012; this is Leahy's formulation):

> **Maximize Presupposition (Leahy)**
> If B is a presuppositional alternative of A, and A and B are merely potential contextual equivalents in c, and π_B is stronger than π_A, utter B.

A and B are merely potential contextual equivalents in c iff A and B have identical assertive contents in the context $c' \neq c$ that results from accommodating any presuppositions of A/B not satisfied in c. The implicature arises by the fact that the speaker chose to utter a sentence whose presuppositions are weaker than one of its presuppositional alternatives—by failing to convey the additional information presupposed by B, the speaker implicates that π_B is false.

Consider the following example:

(43) John lost a basil plant last winter.

An utterance of (43) implicates that John had more than one basil plant. This implicature is generated as an antipresupposition as follows. (43) has the following presuppositional alternative:

(44) John lost his basil plant last winter.

(43) presupposes nothing (at least about John's basil plants), whereas (44) presupposes that John had exactly one basil plant. The latter's presupposition is stronger than the former's. Furthermore, (43) and (44) are potential contextual equivalents—in a context in which (44)'s presuppositions are satisfied, the two have identical assertoric content

(roughly, there is a basil plant that John lost last winter). Yet the speaker chose to utter (43). Assuming that the speaker is following Maximize Presupposition, then, we infer that she does not believe (44)'s stronger presupposition. Assuming furthermore that she is relevantly informed (cf. Sauerland 2004), we then infer that she believes its presupposition to be false.

Leahy extends this strategy to predict counterfactuality implicatures of subjunctive conditionals. The ingredients needed for this explanation to work are as follows:

Leahy's Assumptions

1. An indicative and its subjunctive counterpart are potential contextual equivalents in some contexts.
2. A subjunctive has its indicative counterpart as a presupposition alternative.
3. An indicative presupposes that its antecedent is compatible with the information of the context.
4. A subjunctive presupposes nothing.

From these assumptions, Leahy is able to derive the counterfactuality implicature of subjunctive conditionals as an antipresupposition, as follows. Suppose I utter:

(16) If Oswald had not shot Kennedy, someone else would have.

By Assumption 2, this has as (15) as a presupposition alternative:

(15) If Oswald did not shoot Kennedy, someone else did.

Furthermore, by Assumption 3 and Assumption 4, the presupposition of (15) is stronger than that of (16). Thus, given Assumption 1 and assuming the speaker uttering (16) is obeying Maximize Presupposition, we will infer that the speaker does not believe (15)'s stronger presupposition, and thus that she believes it is false; thus, we infer that the speaker believes that Oswald shot Kennedy.

Leahy's theory is quite promising. That said, I have several worries about it. First, we have reason to doubt Assumption 1—as we saw in §8.2.3, there is independent reason to think indicatives and subjunctives are not truth conditionally equivalent. And although you cannot believe a metaphysical subjunctive without also believing its epistemic indicative counterpart, you *can* believe an epistemic indicative without also believing its metaphysical subjunctive counterpart.

Perhaps we could find some alternative explanation of these observations, if Leahy's story is otherwise promising and independently motivated. However, there are reasons to think that the antipresupposition account of sentences like (43) should in principle not extend to conditionals. The standard motivation for antipresuppositions is that they promise an explanation of the oddity of sentences like (see in particular Heim 1991):

(45) #John broke all of his arms.

According to the standard antipresupposition story, (45) is odd because there is a presuppositionally stronger alternative (*John broke both of his arms*) that carries the same assertoric information as (45)—in such a case, by the standard formulation of Maximize Presupposition, one ought to utter the latter rather than (45):

> **Maximize Presupposition (Standard)**
> If B is a presuppositional alternative of A, the presuppositions of each are satisfied in c, each have the same assertoric content in c, and π_B is stronger than π_A, utter B.

	Presupposes	Asserts
Both Fs are Gs	There are exactly two *F*s	Every *F* is a *G*.
All Fs are Gs	There is at least one *F*	Every *F* is a *G*.

In this context, *both* is a presupposition alternative of *all*, the presuppositions of each are satisfied in the context (since it is presupposed that John has exactly two arms), and thus in that context the assertoric contents of the two are the same; yet, *both* carries a stronger presupposition than *all*. So, we predict that (45) should be odd—the speaker should have said *John broke both of his arms* instead.

The dilemma for extending the antipresupposition story to conditionals is this. In an exactly analogous scenario, uttering the conditional with the weaker presupposition is not infelicitous. The case I have in mind is this: we are not sure whether the butler committed the murder. I say,

(46) If the butler had done it, he would have used a knife. (So let us find out what the murder weapon was.)

	Presupposes	Asserts
$B \rightarrow K$	$\Diamond_E B$	All relevant B-worlds are K-worlds.
$B \Rightarrow K$	Nothing	All relevant B-worlds are K-worlds.

(Remember, $B \rightarrow K$ is the indicative and $B \Rightarrow K$ the subjunctive.)

In this case, the indicative counterpart of (46)'s presupposition is satisfied in the context, the two (according to Leahy's assumptions) have the same assertoric content, and the indicative's presupposition is stronger than the subjunctive's. So, by Maximize Presupposition (Standard), the subjunctive should be infelicitous in this context, just as (45) was above.

This is why Leahy reformulates the maximize presupposition principle as Maximize Presupposition (Leahy). His principle, by contrast, predicts that in the case just described, (46) is acceptable. This is because (47) and (46) are not merely potential contextual equivalents—in the context described, since the presuppositions of both are satisfied, they are contextual equivalents.

(47) If the butler did it, he used a knife.

However, Leahy's principle then does not predict that (45) is odd. Again, in that context, it is presupposed that John has exactly two arms, in which case *both* and *all* are not merely potentially contextually equivalent—their presuppositions are both satisfied so they are contextually equivalent. But then we do not predict that the speaker ought to say *both* rather than *all*, and hence we do not predict that

(45) is infelicitous. Instead, it seems that saying *all* leads us to wonder whether it really is common ground in the context that John has exactly two arms.

The dilemma is this. To capture the infelicity of (45) in the normal context, Maximize Presupposition (Leahy) won't do—we need something like Maximize Presupposition (Standard). But to capture the fact that (46) does not implicate the falsity of its antecedent, we need Leahy's principle, not the standard one. So, we have reason to think that the explanation of the counterfactuality implicature will be different from the explanation of (43) and (45).

8.4.2 Looking Elsewhere

The contrast illustrated at the end of the last section shows that the counterfactual implicature of subjunctive conditionals is more fickle than an antipresupposition—it is not present in contexts where we would expect an antipresupposition to arise. It seems we need to look elsewhere for an adequate theory of the counterfactuality implicature. In this section, I lay out a rationale that explains why we, by default, infer that a speaker uttering a subjunctive conditional believes that its antecedent is false.

I want to start not with antecedents, but consequents. As we will see, reasoning about the speaker's beliefs about the consequent may be a more fruitful path to inferring that the speaker believes its antecedent is false. Here is the reasoning that gets us, by default, from the fact that the speaker uttered a subjunctive conditional to the conclusion that the speaker believes its antecedent is false. Throughout, let '$A \Rightarrow B$' denote an arbitrary subjunctive conditional that the speaker has just asserted.

- Step 1: Suppose the speaker believes **B**.
  - Then, if they also believed **A**, they would believe **AB**, which (by Strong Centering) logically entails the subjunctive $A \Rightarrow B$ (but not vice versa). Thus, in that case, they should have just said $A \wedge B$, which is more informative than $A \Rightarrow B$. So, we conclude that the speaker does not believe **A**.

- Step 2: Suppose the speaker does not believe **B**.
  - Then, if they also believed **A**, there would be some **A$\bar{\mathbf{B}}$**-worlds compatible with their beliefs. But every such world is one at which $A \Rightarrow B$ is false (by Modus Ponens; we are restricting attention to simple subjunctives here). So, since we assume the speaker believes what she says, we conclude that the speaker does not believe **A**.

Conclusion 1: So, the speaker does not believe **A**.

- Step 3: We default assume that the speaker is opinionated about whether **A**—that is, they either believe **A** or believe $\bar{\mathbf{A}}$ (cf. Soames 1982, Horn 1989, Zimmerman 2000, Sauerland 2004, Geurts 2005, Van Rooij & Schulz 2004).

Conclusion 2: So, the speaker believes $\bar{\mathbf{A}}$.

The crucial default assumption that may be overruled is Step 3. This assumption is standardly appealed to in deriving scalar implicatures (see the citations above). However, since this assumption can be overruled, we predict the possibility of non-counterfactual subjunctives like (46):

(46) If the butler had done it, he would have used a knife. (So, let us find out what the murder weapon was.)

We can also predict the possibility of Anderson-style non-counterfactual subjunctives, for example:

(40) If Jones had taken arsenic, he would have shown just exactly those symptoms that he does in fact show.

To simplify the case, suppose we know Jones is showing symptom X. Then, (40) is equivalent in our context to:

(48) If Jones had taken arsenic, he would have exhibited symptom X.

The speaker believes (48) and that Jones is exhibiting symptom X. Here, we block the reasoning at Step 1, because even though Jones took arsenic and is exhibiting symptom X is logically stronger than (48), the conditional is more relevantly informative here because it reveals the speaker's basis for believing that Jones took arsenic: namely, that arsenic is a cause of symptom X.[9]

Summary: *Counterfactuality*

I argued that the counterfactuality implicature of subjunctive conditionals is not an anti-presupposition; instead, I hold instead that it is a standard Quantity implicature, derived on reasoning about the speaker's beliefs about the antecedent.

8.5 Summary

In this chapter I defended the following thesis about the tense morphology of subjunctive conditionals:

> **Temporal Past**
>
> The time of a subjunctive conditional (at which its covert modal's domain is determined) is to the past of utterance time. The time of an conditional is utterance time.

I argued that this thesis, together with some independent assumptions about modal bases (that historical modal bases are responsible for metaphysical interpretations, and that informational modal bases are responsible for epistemic interpretations) and a pragmatic constraint on the assignment of modal bases in context (Avoid Trivialization),

[9] Notice that the same reasoning gets us to Step 3 for indicative conditionals as well. However, the opinionated assumption in Step 3 is suspended for indicative conditionals, since we could then infer at that step that the speaker asserting $A \rightarrow B$ believed A is false, but this is incompatible with the presupposition of $A \rightarrow B$ (namely, that its antecedent is epistemically possible).

is sufficient to predict the range of modal flavors that we in fact find for indicative and subjunctive conditionals, as well as generate the counterfactual implicature in the latter.

However, the resulting theory seems to generate very unusual truth conditions for metaphysical conditionals. A conditional like

(49) If I had dropped this ball, it would have fallen.

is predicted to be (determinately) true if and only if every world that shares our history up until some time before I chose not to drop the ball, at which I did drop the ball, is one at which it fell. But there will be a great many such worlds, some of which obey our laws and others that violate them. Thus, there will be worlds in that set at which I drop the ball and it falls, and worlds at which I drop the ball and gravity inverts and it hangs in mid-air. So, we predict that (49) is not determinately true, contrary to my intuition at least. And so our predicted truth conditions are too weak, it seems. In the next chapter, I discuss my strategy for predicting more plausible truth conditions for metaphysical conditionals.

Appendix

Morphosemantics

Recall the observation from Chapter 6 that subjunctive conditionals contain fake past tense morphology:

(50) I thought the flight was tomorrow, not today! If the plane had left tomorrow, I would have caught it.

There really are two observations here (as pointed out by Mackay 2017). The first is that, unlike past indicatives, whose antecedents and consequents must be about the past, two-past subjunctive conditionals allow for antecedents and consequents about the past, present, or future:

(51) a. If John was sick $\left\{\begin{array}{l}\text{earlier}\\ \text{\#right now}\\ \text{\#later today}\end{array}\right\}$, he went home early.

b. If John had been sick $\left\{\begin{array}{l}\text{earlier}\\ \text{right now}\\ \text{later today}\end{array}\right\}$, he would have gone home early.

The second is that one-past subjunctive conditionals, unlike both past indicatives and two-past subjunctives, do not allow their antecedents/consequents to be about the past:[10]

(52) If John were sick $\left\{\begin{array}{l}\text{\#earlier}\\ \text{right now}\\ \text{later today}\end{array}\right\}$, he would go home early.

Summarizing:

Morphosemantic Observation

- Two-past subjunctives allow their antecedents/consequents to be about past/present/future events.
- One-past subjunctives allow their antecedents/consequents to be about present/future events only.
- Past indicatives allow their antecedents/consequents to be about past events only.

[10] This observation is more complicated than this, but I lack the space to examine the issue further. As observed by Iatridou 2000: 250–251, important differences arise for telic and individual/stage level stative predicates:

(i) If John danced . . . Telic: Dance time after utterance time

(ii) If John owned a cat . . . Individual stative: cat-ownership-time overlaps utterance time

(iii) If John were tired . . . Stage level stative: tired-time may overlap or be after utterance time

This observation raises a prima facie problem for both modal past and temporal past views. Modal past theories nicely explain what is going on with one-past subjunctives. There, the layer of past tense contributes only modal distancing, and so we expect one-past subjunctives to behave like simple present sentences: they can be about either the present or the future. Then, the modal past theorist can hold that two-past subjunctives involve two layers of past tense—one interpreted modally and one interpreted temporally. But if the second layer of past tense is interpreted temporally on antecedent/consequent, we expect these clauses to behave like simple past-tensed sentences, in which case we predict that they must be about the past. Thus, on this view, we don't predict why two-past subjunctives allow, but don't require, that their antecedents/consequents to be about the past.

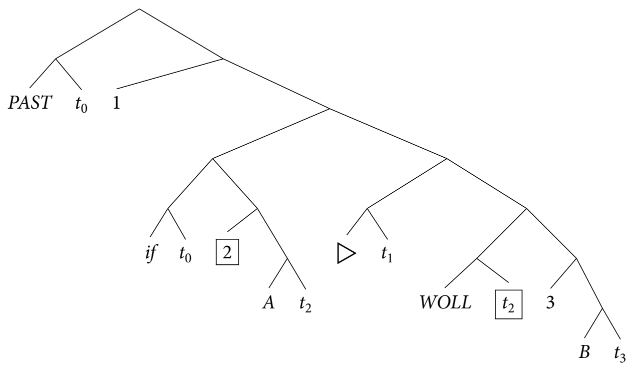
By contrast, a temporal past view can explain what is going on with two-past subjunctives. For the temporal past theorist, the past tense scopes above the conditional, and the morphological marker on the antecedent/consequent is a reflex there to indicate agreement with the higher tense (in other words, it is semantically null)—that's why two-past subjunctives can be about the past, present, or future. But the temporal past view has trouble explaining why one-past subjunctives behave differently. If the same tense configuration holds for one-past subjunctives, then we predict that they should likewise be able to be about past times.

So, it seems, the Morphosemantic Observation raises trouble for both modal and temporal past theories. As such, I suspect they may share a solution to this problem, although I won't explore that point further. Rather, in what follows, I will sketch an outline of a solution to this issue on behalf of the temporal past theory. I will not attempt to implement this strategy compositionally—that would require significant more space and time than I have here (though see Romero 2014, Schulz 2014 for compositional strategies in this area).

There are several ideas in my proposal. The first is that all subjunctive conditionals contain a layer of past tense that scopes above the entire conditional and shifts the evaluation time of its modal to the past (this is the primary commitment of a temporal past view). The second is that the past tense morphology appearing in one- and two-past conditionals is null—it merely marks agreement with the higher past tense. The third is that *if*-clauses contain an indexical temporal element: they shift the evaluation time of their complement to be either identical or after utterance time (cf. Kaufmann 2005b).[11] The fourth is that the untensed future morpheme *woll* is anaphoric to the higher past tense operator, and shifts the evaluation time of its complement to some time overlapping or after that. The fifth is that two-past conditionals additionally have two instances of perfective morphology (*had* and *have*) that I assume is anaphoric on a higher temporal operator (*if* in the antecedent and *woll* in the consequent) and shifts the time of evaluation of its complement to the past of that time (perhaps among other things).

Here is the tense structure I am conjecturing:

One-past subjunctive:

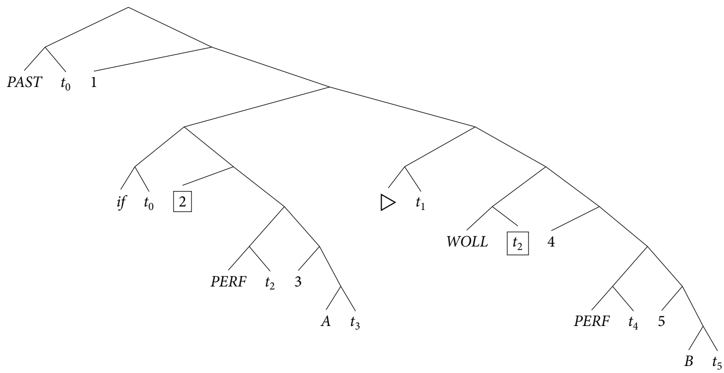


[11] An independent reason for positing such a feature of *if*-clauses comes from the contrast between:

(i) a. ?John comes to school tomorrow.
b. John will come to school tomorrow.
c. If John comes to school tomorrow, he will bring his lunch with him.

The (a)-sentence is odd, except on its scheduling reading (Dowty 1979, Comrie 1985, Copley 2002), where it is interpreted as meaning that it is now determined (according to some schedule) that John will come to school tomorrow. By contrast, (b) and the antecedent of (c) have that interpretation as well, but also have a non-scheduling interpretation true just if John does in fact come to school tomorrow. Kaufmann 2005b proposes that the scheduling interpretation is the result of a covert modal quantifying over (a restricted domain of) historical possibilities. He proposes that the non-scheduling reading arises when the time of this modal's domain is shifted to the future (the time of John's going to school), in which case the modal claim ends up equivalent to its prejacent. To achieve this result, both *will* and *if* are proposed to be non-past temporal shifters.

Two-past subjunctive:



Notice that the binding is all standard, except for the instance of the variable t_2 appearing next to *WOLL*. This variable is not bound by the index 2 appearing above, so must be dynamically bound to ensure that the time of the consequent is identical to or in the future of the time of the antecedent in one-past conditionals. All together, this yields the following truth conditions for one- and two-past subjunctives (the material under the braces is there to indicate what contributes the relevant temporal shifting, and the time subscripts on **A**/**B** indicate the times they are about):

(53) One-past $A \rightarrow B$ true at m, g, t, w, f iff
$$\underbrace{\exists t_1 < t_0}_{PAST}\underbrace{\exists t_2 \geq t_0}_{if}\underbrace{\exists t_3 \geq t_2}_{WOLL} : f(w, D^{\mathbf{A}^{t_2}}_{w,t_1,m,g}) \in \mathbf{B}^{t_3}$$

(54) Two-past $A \rightarrow B$ true at m, g, t, w, f iff
$$\underbrace{\exists t_1 < t_0}_{PAST}\underbrace{\exists t_2 \geq t_0}_{if}\underbrace{\exists t_3 < t_2}_{PERF}\underbrace{\exists t_4 \geq t_3}_{WOLL}\underbrace{\exists t_5 < t_4}_{PERF} : f(w, D^{\mathbf{A}^{t_3}}_{w,t_1,m,g}) \in \mathbf{B}_{t_5}$$

Again, to emphasize, this is just a statement of some truth conditions, plus a specification for how the tense structure might work to generate them. I haven't yet provided semantic entries for the various units of meaning, nor have I explained how this tense structure is generated from the surface form. My claim is weak: it is just that there is a possible way of generating truth conditions for one- and two-past subjunctives within the temporal past framework that fit the range of possible readings we find for the temporal orientations of the antecedents and consequents of one- and two-past conditionals.

Let's convince ourselves of this weak point. Notice that we predict that, for two-past subjunctives, antecedent and consequent may be in any temporal relation, and bear any temporal relation to utterance time t_u. The time of antecedent t_2 just has to be to the past of some time not to the past of t_u, but this could be any time. And the time of the consequent t_4 just has to be to the past of some time t_3 not to the past of some time t_2 which is to the past of a time t_1 that is not in the past of t_u. But again that could be any time.

More interestingly, we predict that, for one-past subjunctives, neither antecedent nor consequent can be about the past. This is because the time of the antecedent t_1 must not be to the past of t_u, as required by the temporal shifting induced by *if*. And the time of the consequent t_2 must also not be to the past of t_1, which must not be to the past of t_u. Thus, it is possible for a temporal past view to predict, using resources that seem independently reasonably plausible, the range of data we find above.

However, Mackay 2017 considers and rejects a related strategy for the temporal past theorist. I will close with a brief discussion of how my strategy here is slightly different from the ones Mackay considers and rejects. Mackay argues against a related view, in which a present tense operator appears adjacent to the *if*-clause of one-past subjunctives. This operator shifts the evaluation time of the antecedent to either be identical to or after utterance time, which is what is needed. However, Mackay argues that this is ad hoc: in other instances of sequence of tense, there is not an active present tense scoping under the higher past tense.

It should be clear why the proposal here does not rely on positing an active present tense operator in the *if*-clause (and thus underneath the higher past tense operator). Rather, the proposal is that *if* itself contributes the temporal shifting. As far as I can tell, this avoids Mackay's worry, and, furthermore, it is independently plausible—see the discussion of Kaufmann 2005b in footnote 11.

9
Sufficiency Networks

In the previous chapter, I motivated a temporal past theory of subjunctive conditionals, aiming to explain why the morphological differences between indicative and subjunctive conditionals (namely, the presence of fake tense on the latter) generate their corresponding semantic differences (epistemic vs. metaphysical interpretations) and pragmatic differences (openness presupposition vs. counterfactual implicature). In this chapter, I want to refine that theory, focusing on predicting more plausible truth conditions for metaphysical conditionals. According to the theory, metaphysical conditionals express facts about sufficiency networks—causal/explanatory/priority relations between propositions that are determined by the world. This is in stark contrast to informational conditionals, which express purely subjective, epistemic relations between propositions that are relativized to an information state.

9.1 Goodman's Puzzle

The modern literature in analytic philosophy on metaphysical interpretations of subjunctive conditionals begins with Goodman 1947, which presents the following puzzle. Suppose we have a match, which is never struck (and thus never lit), and which is kept dry in a room with ample oxygen. The following seems true:

(1) If the match had been struck, it would have lit.

Goodman's proposal, roughly, is that we accept (1) because there is a consistent set of true propositions that includes the proposition that the match was struck, and that entails, together with the laws, that the match lit. In this case, the set is:

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0010

$$\text{Match}^{+}: \begin{Bmatrix} \text{the match was struck} \\ \text{the match was dry} \\ \text{oxygen was present} \end{Bmatrix}$$

and the relevant law entails the material conditional:

> **Law-1**: (the match was struck and was dry and oxygen was present) ⊃ the match lit

However, immediately after making this proposal, Goodman notices a problem. It seems this very proposal predicts we should also accept (2):

(2) If the match had been struck, it would not have been dry.

After all, the following set of true propositions, together with Law 1, entails that the match was not dry:

$$\text{Match}_{2}^{+}: \begin{Bmatrix} \text{the match was struck} \\ \text{the match did not light} \\ \text{oxygen was present} \end{Bmatrix}$$

But then we should be able to reason to the contradictory of (1) as follows (assuming that there is also a relevant law that a wet match that is struck will not light):

(3) a. If the match had been struck, it would not have been dry.
b. And if the match had been struck and was not dry, it would not have lit.

Goodman's problem reveals the need for further constraints regarding the subjunctive conditional domains (that is, what facts we hold fixed when supposing a subjunctive antecedent). A major goal in the philosophical literature on subjunctives since Goodman has been to provide a revealing answer to this question.

David Lewis offered an influential answer: a subjunctive conditional's domain is made up of the most similar worlds in which its antecedent is true, where similarity is evaluated via the following metric (Lewis 1979a):

1. It is of first importance to avoid big, widespread, diverse (quasi-) violations of law.[1]
2. It is of second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
3. It is of third importance to avoid even small, localized, simple (quasi-) violations of law.
4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

Roughly, a subjunctive conditional $A \Rightarrow B$ is true iff all of the most similar A-worlds are B-worlds.[2] In the match case, we thus suppose the match is struck. We consider four kinds of worlds:

w_1: The match is struck, it is dry, oxygen is present, it lights; the history leading up until just before the striking is exactly the same as it actually is and there is a small (quasi-) violation of law that allows the match to be struck.

w_2: The match is struck, but the history leading to this is completely different (this class of worlds may include ones where the match is not dry and in which oxygen is absent).

w_3: The match is struck, is dry, oxygen is present, but after it is struck, there is a (quasi-) violation of law that prevents it from lighting.

w_4: The match is struck, is dry, oxygen is present, but after it is struck, there is a (quasi-) violation of law that prevents it from

[1] In the deterministic setting, violations of laws are counterexamples to them; in the indeterministic setting, quasi-violations of laws are chance outcomes that seem "to conspire to produce a pattern," such as a fair coin landing heads 10^6 times in a row (Lewis 1986b: 60).

[2] This statement of truth conditions presupposes the Limit Assumption, which is that there aren't more and more similar worlds without end—in other words, that there is a unique set of the most similar A-worlds. Lewis's official theory rejects the Limit Assumption (see Lewis 1973b: 19–20). For discussion, see Pollock 1976, Stalnaker 1980, Bennett 2003, Swanson 2011a, Kaufmann 2017.

> lighting, and furthermore, there are other (quasi-) violations of laws that remove all traces (memories, etc.) of its having been struck.

Lewis's system of weights ensures that w_3 is ranked as more similar than w_4 on the grounds that the latter, but not the former, involves a widespread series of (quasi-) violations of law. It also ensures that w_2 is ranked as more similar than w_3 on the grounds that the latter trades a small (quasi-) violation of law for mere approximate match of particular fact, and approximate match of particular fact is worth less than minimizing small (quasi-) violations of law. Finally, it ensures that w_1 is ranked as more similar than w_2 on the grounds that maximizing region of perfect match of particular fact is more valuable than avoiding small (quasi-) violations of law. Thus, Lewis's strategy seems to correctly predict that (1) is true and (2) is false.

However, Lewis's proposal runs into trouble with cases of hindsight like the following (as pointed out by Slote 1978, Edgington 2004, Schaffer 2004):

> **Coin Toss**. Beth is given the opportunity to bet on the outcome of a fair coin toss. She bets tails. Then, she is given the choice to flip the coin, or let Abe flip the coin. She lets Abe flip the coin and it lands heads, so she loses the bet.

Afterwards, we ask Beth how she wishes things would have gone. Compare the following responses:

(4) I wish I had bet heads.

(5) I wish I had tossed the coin.

In this situation, it seems to me that (4) makes sense, but (5) does not. And this is so because (6) is (factually) true, but neither (7) nor (8) are:

(6) If Beth had bet heads, she would have won.

(7) If Beth had tossed the coin, she would have won.

(8) If Beth had tossed the coin, she would not have won.

However, Lewis's strategy does not predict this result. Compare:

w: Beth bet heads, the coin lands heads, Beth wins.
w': Beth bet heads, the coin lands tails, Beth loses.

Both w and w' will involve a small (quasi-) violation of law to make room for Beth to bet heads instead of tails. But it isn't the case that w' involves an additional (quasi-) violation of law to make room for the coin's landing tails, since the coin flip was indeterminate. Finally, both w and w' promote different aspects of approximate match: w secures match in how the coin landed, while w' secures match in the outcome of the bet. So, it seems Lewis's view should predict w and w' are equally similar, in which case not all of the most similar worlds in which Beth bet heads will be worlds in which she wins, and thus that (6) is false.

By contrast, Lewis's theory makes a better prediction regarding (7) and (8). There, he predicts both conditionals are false. Just as with (6), it seems that worlds w'' and w''' should be predicted to be equally similar:

w'': Beth bet tails, Beth flips the coin, the coin lands tails, Beth wins.
w''': Beth bet tails, Beth flips the coin, the coin lands heads, Beth loses.

The challenge for Lewis is to explain what distinguishes (6) from (7)/(8).

But what plausibly does explain the difference between (6) and (7)? The beginnings of an answer seem to lie with the observation that the fact that the coin landed heads is causally independent of whether Beth bet heads, whereas the fact that the coin landed heads is not causally independent of whether Beth flipped it, since Abe flipping the coin is what actually caused it to land heads. Thus, we might propose the following:

> Whether some A-world is more similar to the actual world than another A-world depends only on facts that are causally independent of whether A.

However, this strategy is not available to Lewis, given his commitment to provide a (non-circular) reductive analysis of causation in terms of counterfactual dependence. However, it is available to someone without such reductive ambitions.

In fact, causal modeling theories of subjunctive conditionals are designed to incorporate exactly such causal information in the determinations of subjunctive domains (for discussion, see Pearl 2000, Hiddleston 2005, Rips 2010, Rips & Edwards 2013, Schulz 2011, Briggs 2012, Kaufmann 2013, Snider & Bjorndahl 2015, Santorio 2019). These theories take as their starting point certain causal facts and use those facts to articulate the truth conditions of subjunctive conditionals. The basic idea is simple: in the betting case, Abe flipping the coin is causally sufficient for the coin landing heads (even though it doesn't deterministically cause this outcome), and Beth betting heads together with the coin landing heads is causally sufficient for Beth winning, while Beth betting heads is causally unrelated to how the coin lands. The truth value of (6) depends on whether Beth wins at worlds in which she bets heads and are otherwise like ours regarding facts causally independent of whether Beth bets heads. Since all such worlds are ones in which the coin lands heads, we predict that (6) is true.

Furthermore, an interventionist theory is also well positioned to predict the truth of (1) in the context above:

(1) If the match had been struck, it would have lit.

Here, we look at worlds where the match is struck that are otherwise like ours regarding facts causally independent of whether the match was struck—these facts include that the match was dry and that oxygen was present, but not that the match didn't light. Then, insofar as striking a dry match in the presence of oxygen is sufficient for it lighting, we predict that (1) is true.

Summary: *Goodman's challenge and causal facts*

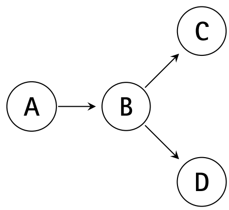
We started our discussion with Goodman's challenge, aiming to explain in a non-circular way what facts we hold fixed when supposing a subjunctive's antecedent. We saw reason to think that Lewis's answer, which appealed to facts about similarity between worlds, fell short. Then we saw some reason to appeal to causal facts in our characterization of what's held fixed in evaluation of subjunctive conditionals—this was the motivating thought behind interventionist theories.

In the next section, I articulate a framework for representing sufficiency networks and sketch the standard interventionist theory of subjunctive conditionals. I will argue that this theory faces several problems, and then argue (in §9.3–9.4) that they are solved by shifting to a temporal past version of the theory that embraces the insights of Chapter 8.

9.2 Sufficiencies and Interventions

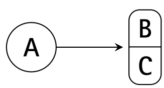
Interventionist theories of conditionals are a subspecies of causal modeling theories. In this section, I will sketch the basic causal modeling apparatus and then discuss the standard interventionist theory. At the core of such a theory are sufficiency relations between propositions. I will use "sufficiency" as a neutral term that encompasses causal, explanatory, or grounding sufficiency—later, in §9.6, we will see how to incorporate each of these into a multi-dimensional theory of subjunctive conditionals. For now, I will focus on causal sufficiency, and take the notion as a primitive, except to say that it is antireflexive (nothing is sufficient for itself) and transitive (if **A** is sufficient for **B** and **B** sufficient for **C**, then **A** is sufficient for **C**), and thus also antisymmetric (if **A** is sufficient for **B**, then **B** is not sufficient for **A**). We will get our grip on the notion of a sufficiency by what it reveals about the meanings of subjunctive conditionals, and vice versa—thus, I am here offering no reductive analysis of either sufficiencies or subjunctive conditionals.

To illustrate the kind of sufficiency facts that metaphysical conditionals express, it will be helpful to make use of the neuron diagrams of Lewis 1986a and causal models of Pearl 2000. A network of sufficiencies relating various propositions can be represented by a directed acyclic graph: roughly, a strict partial order on propositions that represents which are sufficient for which. Here is a sufficiency network:

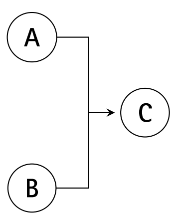


This diagram encodes that **A** is sufficient for **B**, which is sufficient for both **C** and **D**.

As a shorthand, when **A** is sufficient for **B**, we will write '$[\mathbf{A}]\mapsto \mathbf{B}$'. When a proposition is sufficient for a disjunction, as in $[\mathbf{A}]\mapsto (\mathbf{B} \vee \mathbf{C})$, but not sufficient for either disjunct, I will represent that graphically as follows:

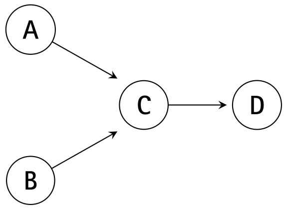


We also have sufficiencies in which a set of propositions are jointly sufficient for some proposition. Where **A** and **B** are jointly sufficient for **C**, we write '$\begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix}\mapsto \mathbf{C}$' and represent this graphically as:



We also distinguish the sufficiencies from the unique sufficiencies within a network. **A** is uniquely sufficient for **B** iff $[\mathbf{A}]\mathrel{\vert\!\!\twoheadrightarrow} \mathbf{B}$ iff **A** is sufficient for **B** and no other proposition is sufficient for **B**.

Finally, sufficiencies encode truth-conditional information that constrains the possible assignments of truth values to propositions: $[\mathbf{A}]\!\rightarrow \mathbf{B}$ encodes $\mathbf{A} \supset \mathbf{B}$; $[\mathbf{A}]\!\twoheadrightarrow \mathbf{B}$ encodes $\mathbf{A} \equiv \mathbf{B}$; and $\begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix}\!\rightarrow \mathbf{C}$ encodes $(\mathbf{A} \wedge \mathbf{B}) \supset \mathbf{C}$. Thus, in the following network of sufficiencies:



$$[\mathbf{A}]\!\rightarrow \mathbf{C},\ [\mathbf{B}]\!\rightarrow \mathbf{C},\ [\mathbf{A} \vee \mathbf{B}]\!\twoheadrightarrow \mathbf{C},\ [\mathbf{C}]\!\twoheadrightarrow \mathbf{D}$$

we have the following possible assignments of truth values:

A	**B**	**C**	**D**
T	T	T	T
F	T	T	T
T	F	T	T
F	F	F	F

Call an assignment of truth values to the propositions in a network of sufficiencies a **sufficiency model**, $\mathcal{M}$. We can formalize a sufficiency model as a set of exogenous propositions and sufficiencies. The **exogenous** propositions in a sufficiency network are those that are not entailed true or false by the consequent of any sufficiency in that network—in other words, they get their truth value from outside of the network, like the setting of the 'initial conditions.' By contrast, the **endogenous** propositions are those that are entailed true or false by the consequent of some sufficiency in that network. In the network above, **A** and **B** are exogenous and **C** and **D** are endogenous. We have the following sufficiency models, corresponding to the four lines in the truth table above:

- $\mathcal{M}_1 = \{\mathbf{A}, \mathbf{B}, [\mathbf{A}]\!\rightarrow \mathbf{C}, [\mathbf{B}]\!\rightarrow \mathbf{C}, [\mathbf{A} \vee \mathbf{B}]\!\twoheadrightarrow \mathbf{C}, [\mathbf{C}]\!\twoheadrightarrow \mathbf{D}\}$
- $\mathcal{M}_2 = \{\mathbf{A}, \bar{\mathbf{B}}, [\mathbf{A}]\!\rightarrow \mathbf{C}, [\mathbf{B}]\!\rightarrow \mathbf{C}, [\mathbf{A} \vee \mathbf{B}]\!\twoheadrightarrow \mathbf{C}, [\mathbf{C}]\!\twoheadrightarrow \mathbf{D}\}$
- $\mathcal{M}_3 = \{\bar{\mathbf{A}}, \mathbf{B}, [\mathbf{A}]\!\rightarrow \mathbf{C}, [\mathbf{B}]\!\rightarrow \mathbf{C}, [\mathbf{A} \vee \mathbf{B}]\!\twoheadrightarrow \mathbf{C}, [\mathbf{C}]\!\twoheadrightarrow \mathbf{D}\}$
- $\mathcal{M}_4 = \{\bar{\mathbf{A}}, \bar{\mathbf{B}}, [\mathbf{A}]\!\rightarrow \mathbf{C}, [\mathbf{B}]\!\rightarrow \mathbf{C}, [\mathbf{A} \vee \mathbf{B}]\!\twoheadrightarrow \mathbf{C}, [\mathbf{C}]\!\twoheadrightarrow \mathbf{D}\}$

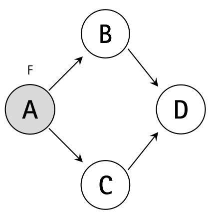
Next, we define an **intervention** into a sufficiency model. An intervention sets the value of some proposition in a sufficiency network, turning it into an exogenous proposition (if it wasn't already) by deleting any sufficiencies that decide the truth of that proposition:

> A sufficiency $[\mathbf{X}] \mapsto \mathbf{Y}$ decides $\mathbf{Z}$ iff either $\mathbf{Y} \vDash \mathbf{Z}$ or $\mathbf{Y} \vDash \bar{\mathbf{Z}}$.

An intervention also deletes any exogenous propositions immediately incompatible with the intervened-on proposition to ensure the intervention is consistent. Thus:

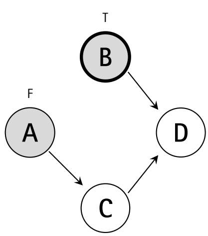
> **Intervention**
> The $\mathbf{X}$-intervention on $\mathcal{M}$ is $\mathcal{M}^{|X}$, the set containing $\mathbf{X}$, every exogenous proposition in $\mathcal{M}$ that is compatible with $\mathbf{X}$, and all and only those sufficiencies in $\mathcal{M}$ that do not decide X.

As an example, take a model like (note that I will shade and label the truth values of exogenous propositions in the network):



$$\mathcal{M} = \{\bar{\mathbf{A}}, [\mathbf{A}] \mapstochar\twoheadrightarrow \mathbf{B}, [\mathbf{A}] \mapstochar\twoheadrightarrow \mathbf{C}, [\mathbf{B}] \mapsto \mathbf{D}, [\mathbf{C}] \mapsto \mathbf{D}, [\mathbf{B} \vee \mathbf{C}] \mapstochar\twoheadrightarrow \mathbf{D}\}$$

Now consider an intervention that sets $\mathbf{B}$, which is false in the original model, to true:



$$\mathcal{M}^{|\mathbf{B}} = \{\bar{\mathbf{A}}, \mathbf{B}, [\mathbf{A}] \mapstochar\twoheadrightarrow \mathbf{C}, [\mathbf{B}] \mapsto \mathbf{D}, [\mathbf{C}] \mapsto \mathbf{D}, [\mathbf{B} \vee \mathbf{C}] \mapstochar\twoheadrightarrow \mathbf{D}\}$$

This **B**-intervened model deletes the sufficiency $[\mathbf{A}] \twoheadrightarrow \mathbf{B}$, turning **B** into an exogenous proposition and setting its value to true. The resulting model contains all of the other sufficiences of $\mathcal{M}$ (since none decide **B**), and so entails $\bar{\mathbf{C}}$ and **D**.

With these definitions in hand, we can now state the standard interventionist account of subjunctive conditional domains:

Interventionism
The domain of $A \Rightarrow B$ at w is the set of worlds whose causal model is $\mathcal{M}_w^{|\mathbf{A}}$, where $\mathcal{M}_w$ is the causal model of w.

Setting aside Strong Centering for a moment, recall that we have:

$$A \Rightarrow B \text{ is } \begin{cases} \text{factually true at } w & \text{if all } \mathbf{A}\text{-worlds in its domain} \\ & \text{at } w \text{ are } \mathbf{B}\text{-worlds} \\ \text{factually false at } w & \text{if all } \mathbf{A}\text{-worlds in its domain} \\ & \text{at } w \text{ are } \bar{\mathbf{B}}\text{-worlds} \\ \text{non-factual at } w & \text{otherwise} \end{cases}$$

Supposing a simple semantics for conditionals for the moment, whereby $A \Rightarrow B$ is true at w iff all of the worlds in its domain are **B**-worlds, we thus generate the following predictions.[3]

Goodman's Match

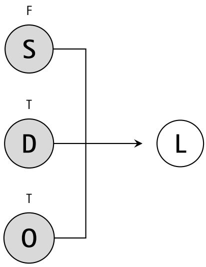
(1) If the match had been struck, it would have lit.

- The context suggests the following sufficiency model:

[3] We could instead have combined this view with my favored semantics, generating slightly different predictions, in particular that (7) is non-factual.

(7) If Beth had flipped the coin, she would have won.

This difference in predictions won't matter for our immediate purposes, so I ignore it for now. Note that within a deterministic model, we would predict ignorance rather than nonfactuality. We will come back to this point when we discuss the probabilities of subjunctive conditionals in Chapter 10.



$$\mathcal{M} = \{\bar{\mathbf{S}}, \mathbf{D}, \mathbf{O}, \left[\begin{smallmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{smallmatrix}\right] \mathrel{\vert\!\!\twoheadrightarrow} \mathbf{L}\}$$

- Intervening in this model to set **S** to true requires replacing $\bar{\mathbf{S}}$ with **S** and leaving everything else intact:

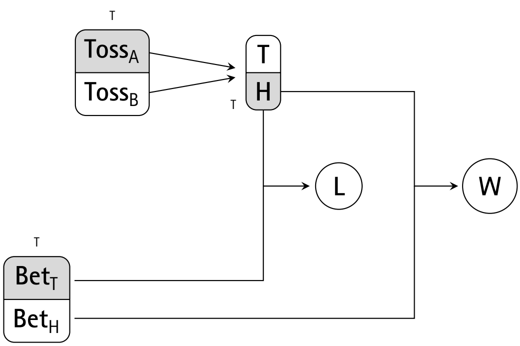
$$\mathcal{M}^{|\mathbf{S}} = \{\mathbf{S}, \mathbf{D}, \mathbf{O}, \left[\begin{smallmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{smallmatrix}\right] \mathrel{\vert\!\!\twoheadrightarrow} \mathbf{L}\}$$

  - And since this set entails **L**, we predict that (1) is true.

Coin Toss

(6) If Beth had bet on heads, she would have won.

- The context suggests the following sufficiency model:



$$\{\ \mathbf{Toss}_A, \mathbf{H}, \mathbf{Bet}_T, \ldots, \left[\begin{smallmatrix}\mathbf{Bet_T}\\ \mathbf{H}\end{smallmatrix}\right] \mapsto \mathbf{L}, \ \ldots\}$$

- Note that we are not representing all the various sufficiencies implicit in the story, in particular, we are not representing $\left[\begin{smallmatrix}\mathbf{Bet_H}\\ \mathbf{T}\end{smallmatrix}\right] \mapsto \mathbf{L}$. Given Intervention, we have:

$$\mathcal{M}^{|\mathbf{Bet_H}} = \{\,\mathbf{Toss_A}, \mathbf{H}, \mathbf{Bet_H}, \begin{bmatrix}\mathbf{Bet_H}\\ \mathbf{H}\end{bmatrix}\!\mapsto \mathbf{W}, \ldots\}$$

And this set entails **W**. Thus, we correctly predict that (6) is true.

However, now notice that, even though $\mathbf{Toss_A}$ is causally responsible for the coin landing heads, since it does not determine this outcome, **H** is an exogenous proposition in this network. As a result, we incorrectly predict that (8) is true:

(8) If Beth had tossed the coin, she wouldn't have won.

Our intervention to set $\mathbf{Toss_B}$ to true forces us to leave **H** true, since it is compatible with $\mathbf{Toss_B}$. But this yields the intervened model:

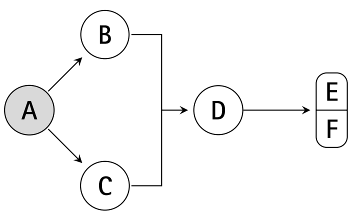
$$\mathcal{M}^{|\mathbf{Toss_B}} = \{\mathbf{Toss_B}, \mathbf{H}, \mathbf{Bet_T}, \begin{bmatrix}\mathbf{Bet_T}\\ \mathbf{H}\end{bmatrix}\!\mapsto \mathbf{L}, \ldots\}$$

And the resulting set entails that Beth didn't win.

The problem is that we should not hold fixed the fact that the coin landed heads when we intervene to set $\mathbf{Toss_B}$ to true, and this is intuitively because Beth's tossing the coin would be a *different causal source* for the heads outcome than Abe's tossing the coin.

9.2.1 Fix: Governance

To incorporate this insight into our theory, we need a way to distinguish exogenous propositions that are governed by some prior proposition from those that are not. The intuitive idea is this: a proposition **P** governs **Q** when there is a sufficiency that connects them. Let's start with an example.



In this network, we have the following instances of direct governance:

- **A** directly governs **B** (by way of $[\mathbf{A}] \twoheadrightarrow \mathbf{B}$).
- **A** directly governs **C** (by way of $[\mathbf{A}] \twoheadrightarrow \mathbf{C}$).
- **B** directly governs **D** (by way of $\begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \twoheadrightarrow \mathbf{D}$).
- **C** directly governs **B** (by way of $\begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \twoheadrightarrow \mathbf{D}$).
- **D** directly governs **E** (by way of $[\mathbf{D}] \twoheadrightarrow (\mathbf{E} \vee \mathbf{F})$).
- **D** directly governs **F** (by way of $[\mathbf{D}] \twoheadrightarrow (\mathbf{E} \vee \mathbf{F})$).

Here is the official definition:

Direct Governance
P directly governs **Q** within network $\mathcal{S}$ when there is a sufficiency in $\mathcal{S}$ whose antecedent is **P** or contains **P** as a conjunct and whose consequent is **Q** or contains **Q** as a disjunct.

In addition, we have various instances of indirect governance that result from the transitive closure of governance relations:

- **A** governs **D** (by way of $[\mathbf{A}] \twoheadrightarrow \mathbf{B}$, $\begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \twoheadrightarrow \mathbf{D}$).
- **A** governs **F** (by way of $[\mathbf{A}] \twoheadrightarrow \mathbf{B}$, $\begin{bmatrix} \mathbf{B} \\ \mathbf{C} \end{bmatrix} \twoheadrightarrow \mathbf{D}$, $[\mathbf{D}] \twoheadrightarrow (\mathbf{E} \vee \mathbf{F})$).
  ...

Intuitively, a proposition governs another within a sufficiency network when you can trace a line left to right through the network that starts from that proposition and ends with the one it governs. The official definition goes by the notion of a sufficiency chain:

Sufficiency Chain
A sufficiency chain within some sufficiency network $\mathcal{S}$ is an n-tuple of propositions, such that for any i, its ith member is the antecedent of (or a conjunct of the antecedent of) some $\mathbf{S} \in \mathcal{S}$ and the i_{+1}th member is the consequent of **S** (or a disjunct of the consequent of **S**).

In the network above, we have four maximal sufficiency chains:

$\langle \mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{E} \rangle$
$\langle \mathbf{A}, \mathbf{B}, \mathbf{D}, \mathbf{F} \rangle$
$\langle \mathbf{A}, \mathbf{C}, \mathbf{D}, \mathbf{E} \rangle$
$\langle \mathbf{A}, \mathbf{C}, \mathbf{D}, \mathbf{F} \rangle$

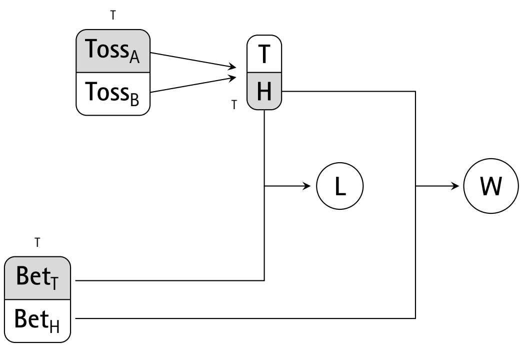
Now, we can define governance formally as follows:

Governance
P governs **Q** in $\mathcal{S}$ iff there **P** comes before $\mathbf{Q}/\bar{\mathbf{Q}}$ in some sufficiency chain in $\mathcal{S}$.

Now, we can modify Intervention to only hold fixed exogenous propositions that are not governed by the intervened-on proposition:

Intervention*
The **X**-intervention on $\mathcal{M}$ is $\mathcal{M}^{|\mathbf{X}}$, the set containing **X**, every exogenous proposition in $\mathcal{M}$ that is compatible with and not governed by **X** (in $\mathcal{M}$), and all and only those sufficiencies in $\mathcal{M}$ that do not decide *X*.

Return now to our sufficiency model for Coin Toss:

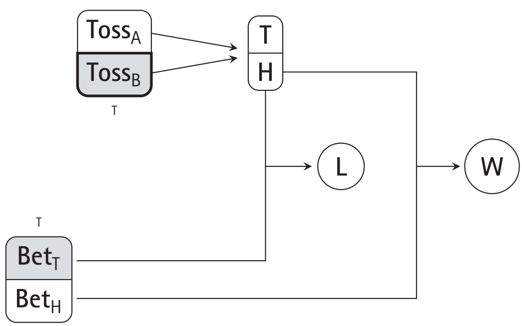


$$\{\ \mathbf{Toss}_A, \mathbf{H}, \mathbf{Bet}_T, \ldots, \begin{bmatrix} \mathbf{Bet_T} \\ \mathbf{H} \end{bmatrix} \mapsto \mathbf{L}, \ \ldots \}$$

Following Intervention*, intervening to set $\mathbf{Toss_B}$ to true does the following:

- Step 1: Eliminate every exogenous proposition that is either incompatible with or governed by $\mathbf{Toss_B}$.
  - Thus, we delete $\mathbf{Toss_A}$ since it is incompatible with $\mathbf{Toss_B}$.
  - And we delete **H** because it is governed by $\mathbf{Toss_B}$ (connected by the sufficiency $[\mathbf{Toss_B}] \mapsto (\mathbf{T} \vee \mathbf{H})$).
- Step 2: Do not eliminate any exogenous proposition compatible with and not governed by $\mathbf{Toss_B}$.
  - Thus, we do not delete $\mathbf{Bet_T}$ because it is compatible with and not governed by $\mathbf{Toss_B}$.
- Step 3: Do not eliminate any sufficiency that does not decide $\mathbf{Toss_B}$.
  - Since none of the sufficiencies decide $\mathbf{Toss_B}$, we hold fixed all of them.

The resulting sufficiency model does not entail either **W** or **L**:



$$\mathcal{M}^{|\mathbf{Toss_B}} = \{\mathbf{Toss_B}, \mathbf{Bet_T}, [\mathbf{Toss_B}] \mapsto (\mathbf{H} \vee \mathbf{T}), \begin{bmatrix} \mathbf{Bet_T} \\ \mathbf{H} \end{bmatrix} \mapsto \mathbf{L}, \begin{bmatrix} \mathbf{Bet_T} \\ \mathbf{T} \end{bmatrix} \mapsto \mathbf{W}, \ldots\}$$

Therefore, we predict that neither (7) nor (8) is true in the context so described. Thus stated, so far, it seems that our interventionist theory of subjunctives is doing quite well. I turn next to raise some challenges to it.

9.3 Challenges to Interventionism

I will raise two problems to Interventionism, one stemming from backwards subjunctives and the other from backtracking subjunctives. These are often not properly distinguished in the literature, so let me explain the difference between them here.

A backwards subjunctive is one whose consequent is about a time that comes before the time its antecedent is about. Here is an example:

(9) If Beth had won, she would have bet on tails.

Backwards subjunctives are contrasted with forward subjunctives, whose consequents are about times equal to or to the future of the times of their antecedents (all of the other subjunctives discussed so far have been forward subjunctives).

Backtracking subjunctives, by contrast, are subjunctives that have a particular, non-standard backtracking interpretation. To illustrate the difference between backtracking and normal subjunctives, here is an example from Lewis 1979a (by way of Downing 1959):

> Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel; if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday. In that case Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all. (Lewis 1979a: 456)

The person making this speech does not contradict himself, even though he utters both:

(10) If Jim asked Jack for help today, Jack would not help him.

(11) If Jim asked Jack for help today, Jack would help him.

The reason is that (10) and (11) are both true, but on different interpretations—(10) is true on its normal interpretation, while (11) is true on its backtracking interpretation. The backtracking interpretation is so-called because it involves backtracking to make certain changes that are necessitated by the antecedent and then letting those changes ramify forward throughout the system. The backtracking interpretation of (11) is plausibly true because backtracking from the supposition that Jim does ask for help today generates a change to how things were before Jim's query—in particular, to thinking that Jim would have asked for help only if there had been no quarrel prior—and given the lack of a prior quarrel, Jack would then have acquiesced to Jim's request.

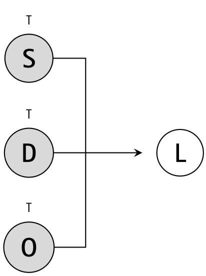
The first problem for Interventionism comes from backwards subjunctives. Consider a variation on Goodman's case from before: this time, suppose the match is struck, dry, and oxygen is present, and thus the match lights. Now consider the following:

(12) a. If the match hadn't lit, it wouldn't have been struck.
b. If the match hadn't lit, it wouldn't have been dry.
c. If the match hadn't lit, there wouldn't have been oxygen present.

It seems to me that none of (12-a)–(12-c) are clearly true (or clearly false). Each is possible, but nothing settles which way things would have to have been had the match not lit. By contrast, Interventionism predicts that each of (12-a)–(12-c) is clearly false, and instead that (13) is true:

(13) If the match hadn't lit, it still would have been struck, dry, and oxygen would have still been present.

To see why, consider the sufficiency model of the situation from before:



$$\mathcal{M} = \{\mathbf{S}, \mathbf{D}, \mathbf{O}, \left[\begin{smallmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{smallmatrix}\right] \twoheadrightarrow \mathbf{L}\}$$

To evaluate (13), we intervene in $\mathcal{M}$ to set $\mathbf{L}$ to false, but otherwise keep every other proposition intact (thus breaking the sufficiency that leads to $\mathbf{L}$), which yields:

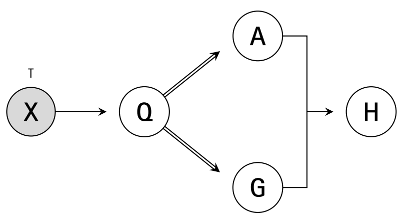
$$\mathcal{M}^{|\bar{\mathbf{L}}} = \{\mathbf{S}, \mathbf{D}, \mathbf{O}, \bar{\mathbf{L}}\}$$

Notice that the sufficiency $\begin{bmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{bmatrix}\twoheadrightarrow \mathbf{L}$ is not among the intervened model. This is because intervening on $\bar{\mathbf{L}}$ requires us to make $\bar{\mathbf{L}}$ an exogenous proposition, and thus delete any sufficiencies whose consequents decide it. However, any world with the resulting sufficiency model is one at which **S, D**, and **O** are all true; hence the view predicts (incorrectly, it seems) that (13) is true.

Next, consider Lewis's backtracking subjunctive (11):

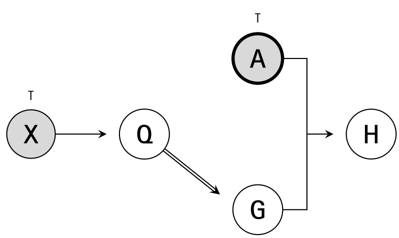
(11) If Jim asked Jack for help today, Jack would help him.

We have something like the following causal model (**X** = the source of Jim and Jack's quarrel, **Q** = Jim and Jack quarreled, **G** = Jack is feeling generous, **A** = Jim asks for help, **H** = Jack helps Jim), using the double arrow to indicate $[\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{A}}$ and $[\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{G}}$:



$$\mathcal{M} = \{\mathbf{X}, [\mathbf{X}]\twoheadrightarrow \mathbf{Q}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}\}$$

Interventionism predicts that in evaluating (11), we intervene on $\mathcal{M}$ to set **A** true:



$$\mathcal{M}^{|\mathbf{A}} = \{\mathbf{X}, \mathbf{A}, [\mathbf{X}] \twoheadrightarrow \mathbf{Q}, [\mathbf{Q}] \twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix} \twoheadrightarrow \mathbf{H}\}$$

However, the resulting model still entails $\bar{\mathbf{H}}$, and thus we predict that (11) is false.

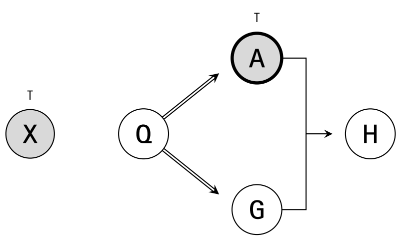
Summary: *Interventionism challenged*

We sketched an interventionist theory of subjunctives according to which, in evaluating $A \Rightarrow B$, we envision an intervention into a sufficiency network just before the antecedent A to set it to be true. This view proved an advance over the earlier, unconstrained similarity theory, but faced challenges stemming from backwards and backtracking subjunctive conditionals.

In the next section, I argue that combining the interventionist insights with the temporal past theory from Chapter 8 provides a solution to these problems.

9.4 The Historical-Sufficiency Theory of Subjunctive Conditionals

Let's step back from Interventionism for a moment to think about the notion of intervening in a sufficiency model in a slightly different way. Notice that, in Lewis's example, we have two potential intervention points where we could break the causal order in order to set **A** to true: one spot is severing the sufficiency between **Q** and $\bar{\mathbf{A}}$ and the other is severing the connection between **X** and **Q**. Breaking the first leads to the sufficiency model above, while breaking the second leads to the following model:



$$\mathcal{M}^{*|\mathbf{A}} = \{\mathbf{X}, \mathbf{A}, [\mathbf{Q}] \twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}] \twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix} \mathbf{A} \\ \mathbf{G} \end{bmatrix} \twoheadrightarrow \mathbf{H}\}$$

Notice that, here, setting **A** to true necessitates letting **A** control the truth value of **Q**, and thus also **G**. The change to **A** thus brings about an upstream change that in turn brings about further downstream changes (now setting **H** to true). And, furthermore, letting this model determine the domain of (11) allows us to predict it is true.

I think that this is exactly how we should think about backtracking interpretations of subjunctive conditionals: they are the result of earlier than default interventions into the sufficiency model.[4] Thus, in order to account for these interpretations, I propose that we distinguish the intervention point from the antecedent; the former is constrained, but not determined by, the latter. Similarly, I think we should say something similar about backwards subjunctives. The intervention point must come before the consequent as well, which means that when the consequent precedes the antecedent, the intervention point will (often) be earlier than when the antecedent precedes the consequent.

Notice that by distinguishing the intervention point from the antecedent and consequent, we have relativized subjunctive conditionals to an additional parameter of evaluation. And notice that this parameter seems suspiciously like a time (note my use of the adjectives *earlier* and *later*). I claim that the intervention point for a subjunctive just is its conditional time. Thus, our reasoning about backwards and backtracking subjunctives provides further motivation for the temporal past theory motivated in Chapter 8.[5]

This is all well and good, but it remains to be seen how to implement these insights in a serious theory. I turn to that issue next.

[4] In Khoo 2017, I proposed a similar account, although did not connect the strategy to similar approaches within causal model theories of subjunctive conditionals.

[5] A similar strategy for handling backwards subjunctives in an interventionist theory is suggested by Santorio 2019, although he does not spell out the proposal in detail.

9.4.1 Building Metaphysical Domains (Informally)

Turn back now to the semantics for subjunctive conditionals motivated in Chapter 8, and consider (6):

(6) If Beth had bet on heads, she would have won.

So far, we have said that the domain of (6) consists of all of the worlds in which Beth bet heads that are like ours until some time right before, and then which may diverge thereafter. But that domain will include worlds where Beth bets heads and the coin lands tails (despite the fact that it actually landed heads). So, we do not yet predict that (6) is (determinately) true in the scenario as described.[6]

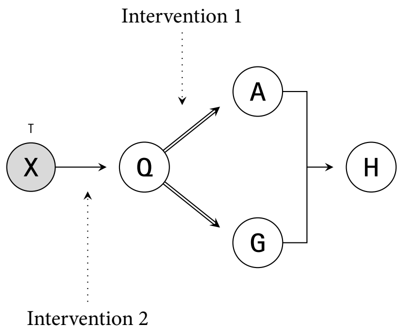
To remedy this point, we need to let facts after the intervention point constrain the domains of metaphysical subjunctive conditionals. We can do this by adding a second component to the determination of subjunctive domains, an ordering source, as in Kratzer 1981, 1991, 2012.

Here is the basic idea. Conditional domains are built up in three steps:

- Step 1: Hold fixed all facts about times up until conditional time (the intervention point).
- Step 2: Hold fixed all sufficiencies about times after conditional time.
- Step 3: Hold fixed all facts about times after conditional time that are independent of (that is, whose values aren't determined by) the antecedent. We will define this notion of independence formally below.

Informally, here is how this recipe works to generate the domains associated with intervening in the two different times of the quarrel sufficiency network:

[6] This is a challenge that von Fintel 2012b raises for a historical modality theory of subjunctive conditionals.



$$\mathcal{M} = \{\mathbf{X}, [\mathbf{X}]\twoheadrightarrow \mathbf{Q}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}\}$$

Suppose we first intervene just before **A** to set **A** to true. Thus:

- Step 1: Hold fixed all the facts prior to this intervention point: **X** and **Q** are thus held fixed.
- Step 2: Hold fixed all of the sufficiency facts about times after this intervention point: $\begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}$.
- Step 3: Hold fixed all facts about times after the intervention point that are independent of the antecedent: this is just $\bar{\mathbf{G}}$.

The resulting domain determined by Intervention 1 is $\{\mathbf{X}, \mathbf{Q}, \mathbf{A}, \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}\}$. And since this set entails $\bar{\mathbf{H}}$, as evaluated at this intervention point, $A \Rightarrow \neg H$ is true.

Suppose now we intervene before **Q** to set **A** to true. Thus:

- Step 1: Hold fixed all the facts prior to this intervention point: **X**.
- Step 2: Hold fixed all of the sufficiency facts about times after this intervention point: $[\mathbf{X}]\twoheadrightarrow \mathbf{Q}$, $[\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{A}}$, $[\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{G}}$, $\begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}$.
- Step 3: Hold fixed all facts about times after the intervention point that are independent of the antecedent: none of the facts meet this criterion.

The resulting domain determined by Intervention 1 is $\{\mathbf{X}, \mathbf{A}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\twoheadrightarrow \mathbf{H}\}$. And since this set entails **H**, as evaluated at this intervention point, $A \Rightarrow H$ is true.

9.4.2 Building Metaphysical Domains (Formally)

That's the informal description of the theory. Now, we turn to state the theory more precisely.

To begin, we modify our historical modal bases to contain all of the historical possibilities up until the relevant time and all of the sufficiency facts about times thereafter, together with the conditional's antecedent.

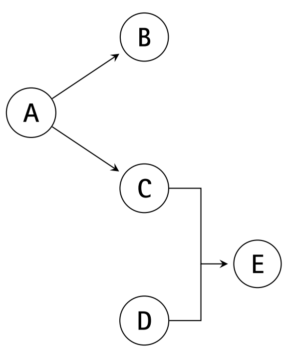
Step 1 (Modal Base).
Historical possibilities up until t + Post-t sufficiency network + **A**.

> **Historical Modality**
> For any w, t and proposition **P**: $\mathbf{P} \in h_{w,t}$ iff **P** is true at w and is about an interval that does not extend beyond t.

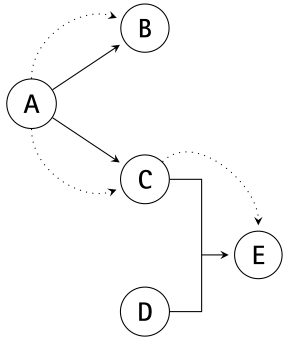
Our ordering source will add to these propositions all and only those post-t truths that are independent of the conditional's antecedent **A**:

Step 2 (Ordering Source).
Post-t facts independent of **A**.

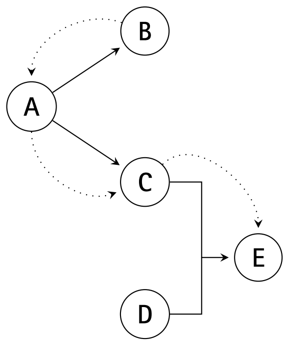
We define independence as follows. Informally, two propositions are **connected** within a network iff you can reach one from the other by tracing through sufficiencies in the network either left-to-right or right-to-left-to-right. Consider the following network:



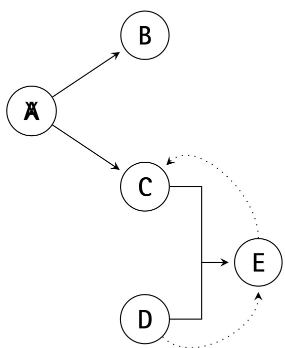
Here, **A** is connected to **B** and **C** and **E**, tracing left-to-right along the sufficiency network:



B is also connected to **A** and **C** and thus **E**, tracing first right-to-left and then left-to-right within the network:



Finally, although **D** is connected to **E** (tracing left-to-right), it is not connected to **C** or any of the other propositions in the network, since this would involve tracing left-to-right and then right-to-left:



Define the notion of two propositions **sharing a root**:

Rootedness
P and **Q** share a root in network $\mathcal{S}$ iff there is an **R** such that **R** governs **P** and **R** governs **Q** in $\mathcal{S}$.

Then, we can formally define what it is for two propositions to be connected in a network:

Connectedness
P and **Q** are connected in network $\mathcal{S}$ iff they appear together in one of its sufficiency chains, or share a root in $\mathcal{S}$.

And we can formally define independence as follows:

Independence
P and **Q** are independent in network $\mathcal{S}$ iff they are not connected in $\mathcal{S}$.

Our metaphysical ordering source g is defined as follows:

Metaphysical Ordering
$g^{\mathbf{A}}_{w,t} = \{\mathbf{P}$: **P** is a post-$t$ truth that is independent of **A** throughout the post-t sufficiency network of $w\}$

Step 3 (Domain).

- A conditional's domain (given h, g) is the set of worlds compatible with the propositions in both its modal base and ordering source:[7]

Conditional Domains
For any w, t, m, g and **A**: $D^{\mathbf{A}}_{m,g,w,t} = \bigcap(h^{\mathbf{A}}_{w,t} \cup g^{\mathbf{A}}_{w,t})$

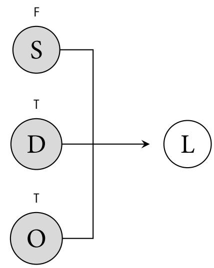
Let's look at some examples to see what this theory predicts. Start with Goodman's original example.

[7] This differs from Kratzer 1981, 1991 since her more complex theory allows for inconsistent ordering sources, whereas the ordering sources we will consider are all consistent.

Case 1: Goodman's Match

(1) If the match had been struck, it would have lit.

<u>Causal model at *w*</u>



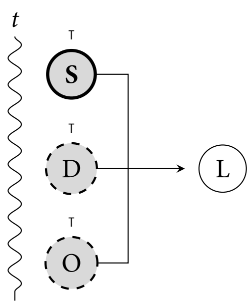
We suppose our conditional time is just t, before **S**. Intervening at t and setting **S** to true yield our modal base, which consists of the propositions about times before t and the post-t sufficiency network:

$$h^{\mathrm{S}}_{w,t} = \{\mathbf{S}, \begin{bmatrix}\mathbf{S}\\\mathbf{D}\\\mathbf{O}\end{bmatrix} \twoheadrightarrow \mathbf{L}\}$$

Our ordering source consists of the post-t truths independent of **S**:

$$g^{S}_{w,t} = \{\mathbf{D}, \mathbf{O}\}$$

since both **D** and **O** are independent of **S**. Notice that $\bar{\mathbf{L}}$ is not in our ordering source, even though it is true at w, since it is not independent of **S**. Thus, we have (I will use a dashed circle to indicate the post-t truths being held fixed in the conditional's domain):



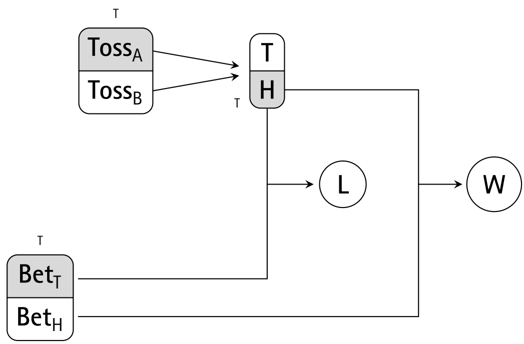
$$D^{\mathbf{S}}_{w,t,h,g} = \bigcap\{\mathbf{S}, \mathbf{D}, \mathbf{O}, \begin{bmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{bmatrix} \twoheadrightarrow \mathbf{L}\}$$

The resulting domain entails **L**, thus predicting the truth of (1).

Case 2: Bets on Indeterminate Coin Tosses

(6) If Beth had bet on heads, she would have won.

Causal model at *w*



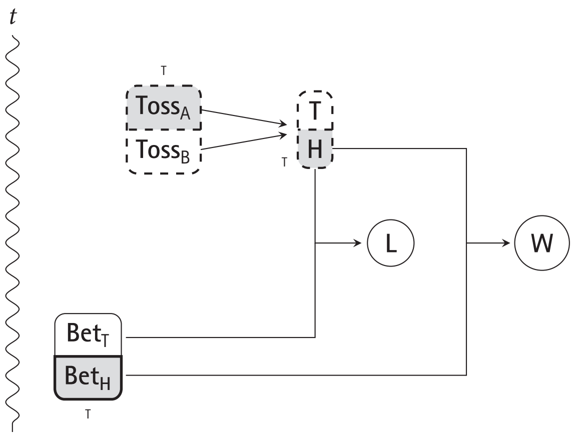
We suppose our conditional time is t, just before Beth bets heads. Then, our historical modal base consists of:

$$h^{\mathbf{Bet_H}}_{w,t} = \{\mathbf{Bet_H}, \begin{bmatrix}\mathbf{Bet_H}\\ \mathbf{H}\end{bmatrix} \mapsto \mathbf{W}, \begin{bmatrix}\mathbf{Bet_H}\\ \mathbf{T}\end{bmatrix} \mapsto \mathbf{L}, \ldots\}$$

And our metaphysical ordering source is:

$$g^{\mathbf{Bet_H}}_{w,t} = \{\mathbf{Toss_A}, \mathbf{H}\}$$

since both $\mathbf{Toss_A}$ and **H** are independent of $\mathbf{Bet_H}$. Notice that **L** is not included because it is not independent of $\mathbf{Bet_H}$, owing to the sufficiency $\begin{bmatrix}\mathbf{Bet_H}\\ \mathbf{T}\end{bmatrix} \mapsto \mathbf{L}$. Thus, our domain is:



$$D^{\mathbf{Bet_H}}_{w,t,h,g} = \bigcap \{\mathbf{Bet_H}, \mathbf{Toss_A}, \mathbf{H}, \begin{bmatrix}\mathbf{Bet_H}\\ \mathbf{H}\end{bmatrix} \mapsto \mathbf{W}, \ldots\}$$

The resulting domain entails **W** and hence that (6) is true. What about (7)/(8)?

(7) If Beth had tossed the coin, she would have won.

(8) If Beth had tossed the coin, she wouldn't have won.

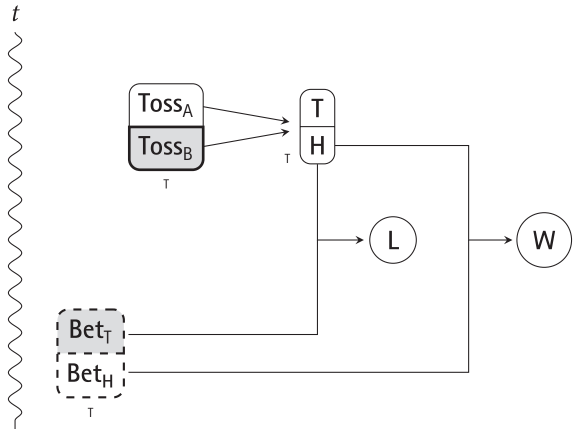
Here, our conditional time is the same, let's suppose. Thus, our modal base is:

$$h^{\mathbf{Toss_B}}_{w,t} = \{\mathbf{Toss_B}, [\mathbf{Toss_B}] \mapsto (\mathbf{H} \vee \mathbf{T}), \begin{bmatrix}\mathbf{Bet_T}\\ \mathbf{T}\end{bmatrix} \mapsto \mathbf{W}, \begin{bmatrix}\mathbf{Bet_T}\\ \mathbf{H}\end{bmatrix} \mapsto \mathbf{L}, \ldots\}$$

And our ordering source is:

$$g^{\mathbf{Toss_B}}_{w,t} = \{\mathbf{Bet_T}\}$$

Notice, in particular, that this ordering source does not contain $\mathbf{Toss_A}$, **H**, or **L**, since each is governed by $\mathbf{Toss_B}$. Thus, we have:



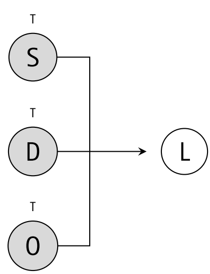
$$D^{\mathbf{Toss_B}}_{w,t,h,g} = \bigcap \{\mathbf{Toss_B}, \mathbf{Bet_T}, [\mathbf{Toss_B}] \mapsto (\mathbf{H} \vee \mathbf{T}), \begin{bmatrix} \mathbf{Bet_T} \\ \mathbf{T} \end{bmatrix} \mapsto \mathbf{W}, \begin{bmatrix} \mathbf{Bet_T} \\ \mathbf{H} \end{bmatrix} \mapsto \mathbf{L} \ldots\}$$

The resulting domain entails neither **W** nor $\bar{\mathbf{W}}$.

Case 3: Goodman's Match Reversed

We now turn to the backwards subjunctive (12-a) (similar remarks apply to the other backwards examples). Remember that in this case, the match is struck while dry and in the presence of oxygen, and lights.

Causal model at *w*



(12) a. If the match hadn't lit, it wouldn't have been struck.

Suppose our conditional time is *t*, just before **S** (this is something I will argue for in §8.4). Our modal base is:

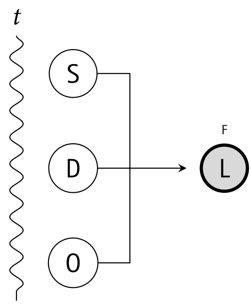
$$h^{\bar{\mathbf{L}}}_{w,t} = \{\bar{\mathbf{L}}, \begin{bmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{bmatrix} \twoheadrightarrow \mathbf{L}\}$$

And our metaphysical ordering source is empty:

$$g^{\bar{\mathbf{L}}}_{w,t} = \{\}$$

Notice that the ordering source does not contain **S**, **D**, or **O** since each governs $\bar{\mathbf{L}}$. So, our domain is:

Causal model at *w*



$$D^{\bar{\mathbf{L}}}_{w,t,m,g} = \bigcap\{\bar{\mathbf{L}}, \begin{bmatrix}\mathbf{S}\\ \mathbf{D}\\ \mathbf{O}\end{bmatrix} \twoheadrightarrow \mathbf{L}\}$$

This domain does not entail $\bar{\mathbf{S}}$; nor does it entail $\bar{\mathbf{D}}$ or $\bar{\mathbf{O}}$. But it does entail their disjunction. Thus, we predict that (12-a) is not factually true. We'll see in Chapter 9 that we predict such conditionals are non-factual, allowing that they may be true.

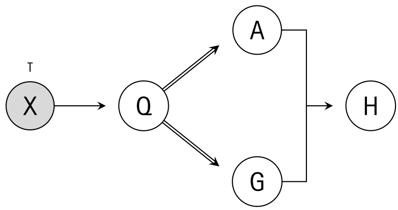
Case 4: Backtracking/Normal Interpretations

Next, I'll show how our theory predicts two interpretations of subjunctives: normal and backtracking. Recall Lewis's example (remember that **X** = the source of Jim and Jack's feud, **Q** = Jim and Jack quarreled, **G** = Jack is feeling generous, **A** = Jim asks for help, **H** = Jack helps Jim):

(10) If Jim asked Jack for help today, Jack would not help him.

(11) If Jim asked Jack for help today, Jack would help him.

Causal model at *w*



$$\text{Sufficiencies} = \{[\mathbf{X}]\!\twoheadrightarrow \mathbf{Q}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\!\twoheadrightarrow \mathbf{H}\}$$

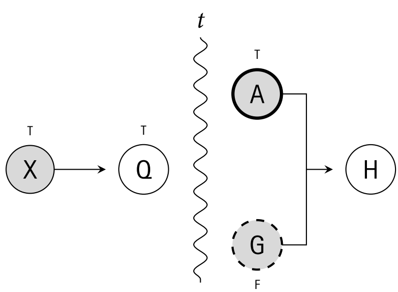
Start with (10), which we aim to predict is true on its normal interpretation. This interpretation is the result of picking a conditional time just before the antecedent **A**; call this time *t*. Our modal base is:

$$h^{\mathbf{A}}_{w,t} = \{\mathbf{A}, \mathbf{X}, \mathbf{Q}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\!\twoheadrightarrow \mathbf{H}\}$$

And our metaphysical ordering source contains $\bar{\mathbf{G}}$, since $\bar{\mathbf{G}}$ is independent of **A**, but doesn't contain either **H** or $\bar{\mathbf{H}}$, since neither is independent of **A**:

$$g^{\mathbf{A}}_{w,t} = \{\bar{\mathbf{G}}\}$$

Thus, our domain is:



$$D^{\mathbf{A}}_{w,t,h,g} = \bigcap\{\mathbf{A}, \mathbf{X}, \mathbf{Q}, \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\!\twoheadrightarrow \mathbf{H}\}$$

The resulting domain entails $\bar{\mathbf{H}}$, and thus we predict that (10) is factually true on this interpretation.

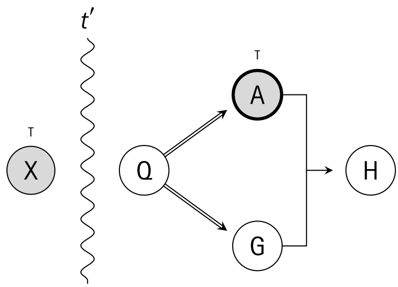
Next, turn to (11), which we aim to predict is true on its backtracking interpretation. This interpretation is the result of picking a conditional time early enough to allow us to backtrack from **A** to **Q**—so our conditional time will be t', just before **Q**. Our modal base is:

$$h^{\mathbf{A}}_{w,t'} = \{\mathbf{A}, \mathbf{X}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\!\twoheadrightarrow \mathbf{H}\}$$

And our metaphysical ordering source is empty:

$$g^{\mathbf{A}}_{w,t'} = \{\}$$

Notice that neither Q nor $\bar{\mathbf{G}}$ is in the ordering source. The former is directly connected to **A** and the latter shares a root with **A**. Thus, our domain is:



$$D^{\mathbf{A}}_{w,t',h,g} = \bigcap\{\mathbf{A}, \mathbf{X}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{A}}, [\mathbf{Q}]\!\twoheadrightarrow \bar{\mathbf{G}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{G}\end{bmatrix}\!\twoheadrightarrow \mathbf{H}\}$$

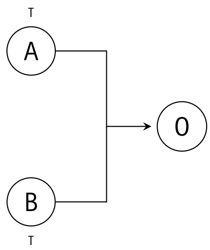
This domain does entail **H**. From **A** and $\mathbf{Q} \equiv \bar{\mathbf{A}}$ we have $\bar{Q}$; and from this and $\mathbf{Q} \equiv \bar{\mathbf{G}}$ we get **G**; and from **A**, **G** and $\mathbf{AG} \equiv \mathbf{H}$ we get **H**. Thus, we predict that (11) is factually true on this interpretation.

Case 5: Non-Factual Backwards Subjunctives

Before we continue, I want to discuss one more type of case to show the value of including our rootedness constraint on metaphysical conditional domains. Consider the following setup:

Light Switches. A light is controlled by two independently operated switches A and B. If and only if both switches are up is the light on. Currently, both switches are up and the light is on.

Here is a model of the scenario:



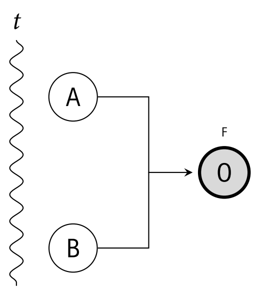
Now contrast the backwards subjunctives:

(14) If the light had been off, **at least** one of the switches would have been off.

(15) If the light had been off, **exactly** one of the switches would have been off.

Intuitively, (14) is true, but by contrast, (15) is not clearly true (nor is it clearly false). Our theory predicts this.

To see why, first we must suppose that the conditional time for each is some time t before the switches have been flipped on (I will argue that our theory can predict this below in §9.5). Thus, our intervened-on model is:



Notice that our modal base contains the post-t sufficiency $\begin{bmatrix}\mathbf{A}\\\mathbf{B}\end{bmatrix} \twoheadrightarrow \mathbf{O}$, but our ordering source does not contain either **A** or **B** because neither is independent of **O**. So, the resulting domain $D^{\bar{\mathbf{O}}}_{m,g,t,w}$ entails $\bar{\mathbf{A}} \cup \bar{\mathbf{B}}$, but is compatible with both possibilities: that both lights are off, or exactly one of the lights is off.

As a result, we predict that (14) is factually true and that (15) is neither factually true nor factually false. Remember, setting aside Strong Centering for now, a conditional is factually true iff its domain entails its consequent, while is non-factual iff its domain entails neither its consequent nor the negation of its consequent:

$$A \rightarrow^{t}_{h,g} B \text{ is } \begin{cases} \text{factually true at } w & \text{if} \quad D^{\mathbf{A}}_{w,t,h,g} \subseteq \mathbf{B} \\ \text{factually false at } w & \text{if} \quad D^{\mathbf{A}}_{w,t,h,g} \subseteq \bar{\mathbf{B}} \\ \text{non-factual at } w & \quad\;\; \text{otherwise} \end{cases}$$

We have seen that the resulting theory makes several nice predictions, including about backwards and backtracking subjunctives.

Summary: *The historical sufficiency theory*

Metaphysical subjunctive conditionals are about a past time (conditional time, or the intervention point) that can vary by context. The intervention point forces us to hold fixed (as a property of the worlds that comprise the conditional's domain) all facts prior to that point, and all sufficiencies about times after that point. In addition, we must hold fixed all post-intervention facts that are independent of the conditional's antecedent. We saw that the resulting theory correctly predicts our intuitions in a wide range of cases, including about forward, backwards, and backtracking subjunctives.

However, many interesting issues remain open. One is what determines the time of a subjunctive conditional. We saw this was crucial to predicting normal vs. backtracking interpretations. But why are normal interpretations the default and what does it take to bring out the backtracking interpretation? I discuss this topic in §9.5. A second issue is the nature of miracles in our theory. On the face of it, sufficiency violations are miracles, in that they involve breaks in the causal order. However, even though we predict miracles in the worlds of a subjunctive's domain, it seems we cannot state that there are such miracles using subjunctive conditionals. In §9.6 I discuss this interesting puzzle. A third issue is whether it is possible to generalize this theory to subjunctives with impossible antecedents. In §9.7 I argue that the theory provides a plausible account of such conditionals that is motivated by a similar asymmetry between forward and backwards subjunctive conditionals.

9.5 What Makes for Backtracking?

Imagine the following situation (Jackson 1977: 9). You come across a crowd of people cheering as a man (call him Smith) exits a tall building, flanked by firefighters. There is a palpable sense of relief among the crowd. You ask what is going on, and an onlooker tells you that a moment ago Smith was positioned on the edge of the roof of the building, poised to jump, but then decided against it and walked down the steps safely. They are relieved because, as they tell you:

(16) If Smith had jumped, he would have died.

As a rough gloss, it seems that this conditional is true just in case, given how things were just before Smith's hypothesized jump, his jumping was sufficient for causing his death. Surveying the scene, you find no nets or other safety devices in place to allow someone to fall safely from that height. In light of this information, it seems that their assessment is correct.

However, now someone else reasons as follows: "Clearly Smith did not want to die—after all, he chose not to jump. Therefore, if he had jumped, there would have been a net beneath him to catch him safely. And thus,

(17) if Smith had jumped, he would have lived."

I find this reasoning also intuitively compelling. Yet, its conclusion seems to contradict something I already accept, namely, (16). Thus, it seems that these conditionals must be true on distinct interpretations: as above, say that (16) is true (and (17) false) on its normal interpretation, whereas (17) is true (and (16) false) on a backtracking interpretation. However, notice that these interpretations are not on a par: while the normal interpretation is the default one, the backtracking interpretation does not seem not available until one uttered the relevant backwards subjunctive (in this case, "if he had jumped, there would have been a net beneath him to catch him safely," or "if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday," in Lewis's example).

Here is another example, based on a variation on the Sly Pete case from Chapter 3 (Gibbard 1981):

> **Sly Pete**. Sly Pete and Mr. Stone are playing poker. Pete is an unscrupulous cheater: he has two of his henchmen, Zack and Jack, planted to help him win. Stone has placed bets on the current hand, and it is now up to Pete to call or fold. Zack has seen Stone's hand, which is quite good, and signaled its contents to Pete. Jack has seen both hands, and seen that Pete has the losing hand. Pete folds.

In light of this information, Pete might feel content with his decision. After all, since he had the losing hand, the following seems true:

(18) If Pete had called, he would have lost.

But, yet, there is a backtracking subjunctive lurking nearby: "Pete knew who had the better hand, and of course he was playing to win. So, if he had called, he would have had the better hand, and thus,

(19) If Pete had called, he would have won."

As before, the normal interpretation seems available by default, with the backtracking interpretation available after some additional contextual preamble.

The generalization seems to be threefold:

(A) Metaphysical subjunctive conditionals allow for backtracking and non-backtracking interpretations.

(B) Normal interpretations tend to be the default.

(C) Accepting a backwards subjunctive sharing the same antecedent as the target conditional helps make salient the backtracking interpretation of the latter.

We have already seen how the theory predicts (A): earlier conditional times yield more backtracking than later conditional times. In the next subsection, I discuss how the pragmatics of conditional time allows us to account for (B) and (C).

9.5.1 The Pragmatics of Conditional Time

We start with two strong constraints on the conditional time of a metaphysical subjunctive conditional:

CT1. Conditional time is before consequent time.

CT2. Conditional time is before antecedent time.

We already motivated CT1 using Avoid Trivialization (see §7.1):

> **Avoid Trivialization**
> If possible, interpret a complex sentence so that it is not epistemically equivalent with any of its sentential constituents.

We can motivate CT2 easily as well, since whenever a metaphysical subjunctive's conditional time is after its antecedent time, then if its antecedent is false, its domain will be empty. Let **A** be the antecedent (and suppose $\bar{\mathbf{A}}$ holds at w, t). Then, by the definition of a historical modal base, for any $t' \geq t$, $h_{w,t'}$ will contain $\bar{\mathbf{A}}$:

> **Historical Modality**
> For any w, t and proposition **P**: $P \in h_{w,t}$ iff **P** is true at w and is about an interval that does not extend beyond t.

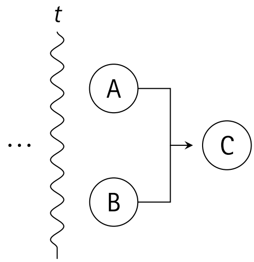
But then $h^{\mathrm{A}}_{w,t'}$ will be inconsistent, and hence $D^{\mathrm{A}}_{w,t',h,g}$ will be empty. Yet, it is a natural assumption that conditionals presuppose that their domains are non-empty—this is a general feature shared by indicative and subjunctive conditionals, as well as by other quantificational expressions. Furthermore, as we saw in Chapter 4, on a sequence semantics we need the conditional's domain to be non-empty for the conditional to be defined. Therefore, it is reasonable to expect that, if possible, conditional time will be some time such that the conditional has a non-empty domain, which yields CT2.

Predicting (B)

Notice that *when* you intervene in a sufficiency network has consequences for how the changes induced by your intervention percolate through the network. In particular, intervening at a time just before **A** will only generate changes in the network that are downstream of **A**, whereas intervening at a time much earlier than **A** will generate changes that are both upstream and downstream of **A**—changes to times before **A** brought about by **A**'s change in truth value, which then have additional downstream consequences. Thus, later conditional times result in an interpretation that encodes information about a more direct sufficiency relationship between antecedent **A** and consequent **C**.

For instance, suppose $A \Rightarrow^t C$ is a forward subjunctive whose conditional time t is just before the time of A. This conditional will encode the information that **A**, together with the history up until **A** and the laws thereafter, along with any truths independent of $\bar{\mathbf{A}}$, is sufficient for **C**. Such a conditional would be true in the following world (suppose that **A** is false, **B** is true, and **C** is false):

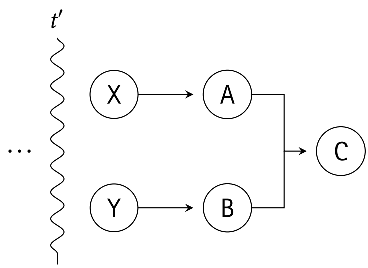
World w_1



Intervening at t to set **A** to true keeps **B** true, and then respecting the post-t sufficiency $\begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix}\!\mapsto \mathbf{C}$ requires setting **C** to true as well.

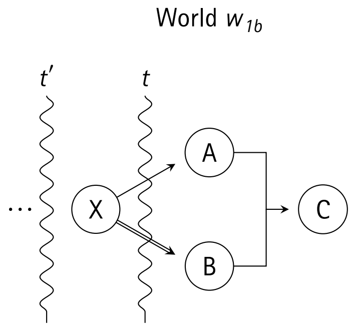
By contrast, a backtracking interpretation of the same conditional (one whose conditional time is $t^- < t$) has us holding fixed less of the actual history and more of its laws, thus changing which post-t facts independent of its antecedent are compatible with its modal base. Let's consider two possible histories to w_1 to show how this difference matters. In w_{1a}, **A** and **B** are governed by independent sufficiencies (here, **A** and **X** are false, and **Y** and **B** are true):

World w_{1a}



We now consider the backtracking conditional $A \Rightarrow^{t'} B$. Intervening at t' to set **A** to true requires backtracking to set **X** true and holding fixed various sufficiencies connecting **X** to **A** and **Y** to **B**. Then, we have two post-t' sufficiencies that are independent of **A**: **Y** and **B**—so, the domain will contain both. Then, respecting the post-t sufficiency $\begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix}\!\mapsto \mathbf{C}$ ensures the domain entails **C**. So, if the history of w_1 is as in w_{1a}, the two interpretations of the conditional $A \Rightarrow B$ will have the same truth value.

But this is not always the case. Sometimes, the upstream consequences of **A** have downstream consequences that help us get to **C**. This is what happens in classic backtracking reasoning. Suppose the history of w_1 is as in w_{1b} instead (here, **X** is false, **A** is false, and **B** is true):



Here, we have $[\mathbf{X}] \twoheadrightarrow \bar{\mathbf{B}}$.

In this world, there is a difference between intervening at t and at t' to set **A** to true. Intervening at t sets **A** to true and keeps **X** false. Since **B** is independent of **A** in the post-t sufficiencies, it will be in the domain of the conditional at t:

$$D^{\mathbf{A}}_{w,t,h,g} = \{\mathbf{A}, \mathbf{X}, \mathbf{B}, \begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix} \twoheadrightarrow \mathbf{C}\}$$

But intervening at t' to set **A** to true has an upstream consequence: this change ensures that **B** shares a root with **A** among the post-t' sufficiencies, and thus **B** is not independent of **A** at t'. Thus, the domain at t' will not contain **B**, and, in fact, its domain entails $\bar{\mathbf{B}}$ and thus $\bar{\mathbf{C}}$:

$$D^{\mathbf{A}}_{w,t',h,g} = \{\mathbf{A}, [\mathbf{X}] \twoheadrightarrow \mathbf{A}, [\mathbf{X}] \twoheadrightarrow \bar{\mathbf{B}}, \begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix} \twoheadrightarrow \mathbf{C}\}$$

Therefore, the backtracking interpretation of $A \Rightarrow B$ is false given this history of w_1.

The lesson we can draw from this is that later conditional times result in interpretations that encode a more direct sufficiency relationship between antecedent and consequent. Earlier conditional times give rise to backtracking interpretations in which upstream and downstream consequences of our intervention matter to the relationship between antecedent and consequent.[8] This is a consequence of requiring that post-t facts that are held fixed must not share a root with the conditional's antecedent in the post-t sufficiencies.

Now, to predict (B), I will argue that information about direct sufficiency relations is generally more relevant to us than information about indirect sufficiency relations.[9] I will consider three pieces of evidence to this effect: direct sufficiency information is relevant to (i) rational decision making, (ii) emotive factive attitudes like regret, and (iii) planning and explanation, but this is not the case for indirect sufficiency information.

(i) *Rational decision making.* A rational deliberator considers how various courses of action available to them will impact the world, given the actual circumstances they find themselves in (cf. Stalnaker 1972, Gibbard & Harper 1981, Lewis 1981). This just is direct sufficiency information connecting various actions with their downstream consequences. By contrast, indirect sufficiency information traces the causal sources of a decision backwards, thus introducing "inadmissible" considerations to the calculus—allowing things outside of your control to influence the consequences of your actions. We find the intuitive contrast here with what DeRose 2010 calls "deliberationally useless" conditionals—those that do not provide instrumental reasons for action. For instance, in the Sly Pete

[8] The same point holds for backwards subjunctives, except here, intervening just before the consequent **C** will result in only downstream effects (from **C**) of intervening to change **A**'s truth value.

[9] Will Starr helpfully points out an alternative hypothesis, which is that backtracking reasoning requires more processing and thus such interpretations may not be cognitively available unless the speaker/hearer is willing to engage in such additional processing, which they do not do so by default.

case, DeRose notes that (20-a) gives Pete a reason not to call, whereas (20-b) does not give Pete a reason to call:

(20) a. If I were to call, I would lose.
     b. If I were to call, I would win.

Pete could rationally reason as follows: I want to win and I have the weaker hand, so if I were to call, I would lose; therefore, I should not call. Pete could not rationally reason as follows: I want to win and I know my opponent's hand, so if I were to call, I would have the stronger hand and thus I would win; so I should call. The latter reasoning mistakenly mixes changes to the past that are not in Pete's control with consequences of his action that are, which is irrational.

(ii) *Emotive factives*. Attitudes like regret and relief require that the attitude-holder believe a counterfactual subjunctive conditional—that things would have been better/worse had the content of the attitude not been true. If I regret telling you that secret, then I must think that things would have been better had I not.[10] But notice that this counterfactual subjunctive must be non-backtracking. Suppose I rank the action/outcome pairs as follows:

- Best: I tell you the secret and you are not sad.
- Second best: I do not tell you the secret and you are not sad.
- Third best: I tell you the secret and you are sad.
- Fourth best: I do not tell you the secret and you are sad.

Aiming for best, I tell you the secret but it makes you sad. I regret my decision because had I not told you the secret, you would not have been sad, which would have been better (second place) than what did happen (third place). But the following backtracking reasoning supporting why I also should not regret my decision makes no sense: had I not told

[10] Notice that the relevant attitude involves a counterfactual comparison with how things went—this is not the same as presupposing that I think my telling you the secret was bad. If all my options were equally bad, then, it seems, I cannot rationally regret doing what I did.

you the secret, that would have been because I knew you were already sad, in which case that would have been worse (fourth best) than what did happen (third best). This seems of a piece with decision making above: regret and other emotive factive attitudes require belief in a non-backtracking counterfactual because it does not make sense to regret an action on the basis of its causal source.

(iii) *Planning and explanation*. We go to the beach and there is a storm. We later learn the non-backtracking interpretation of:

(21) If the wind had changed course, the storm would be delayed until tomorrow.

We thus learn some fact about how the storm's presence causally depended on the wind's direction. This is relevant information for explaining the presence of the storm and for future planning—when in similar circumstances (with a storm nearing), we can rely on this fact and check the wind direction to predict the timing of the storm. But suppose that the wind's direction in this case happened to be the result of a third variable—the temperature of a certain region of the Atlantic Ocean—and that the wind would have only changed course had the temperature been in such a way so as to bring about the storm today anyway. Then, there is a backtracking interpretation of (21) that is false. But this fact is less explanatory, and thus less useful in planning. Why? It seems because it applies to a more limited range of circumstances. It is only applicable in cases where the wind's direction is controlled by a third variable that is hooked up to the storm in exactly the same way. But since we often do not know exactly how these variables are related, we will often not be in a position to apply this fact to planning and explanation in other cases. This point is related to the point above: earlier intervention times allow for both downstream and upstream consequences of reseting **A**'s truth value, and thus for such information to be relevant to planning or explanation, we need to already know enough about the relevant sufficiency network (that is, **A**'s antecedents and any connections they have with factors leading to **C**) to apply that information in our particular case. By contrast, later intervention times only allow for downstream consequences, and this means we can apply such information to any situation in which we know the values of the other factors relevant to **C**.

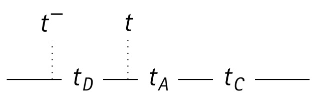
Information about direct sufficiency relations, but not indirect sufficiency relations, matters to decision making, emotive factive attitudes, and planning/explanation. Furthermore, it is hard to see what questions the information expressed by backtracking interpretations could answer, since it seems that we need to already possess the information described by the backtracking interpretation before we access it (this is a point that we will come back to with observation (C), below). That is my explanation for why normal interpretations will typically be the default.

Predicting (C)

We have seen that, by default, a subjunctive will be interpreted with some late admissible conditional time (before the time of its antecedent and consequent). Our final observation was:

(C) Accepting a backwards subjunctive sharing the same antecedent as the target conditional helps with accessing the backtracking interpretation of the latter.

Start with the forward subjunctive whose backtracking interpretation we want to bring out: $A \Rightarrow C$. Since this is a forward subjunctive, t_A is before t_C, so its default interpretation will put its conditional time some time t just before t_A. Now, consider a backwards subjunctive sharing the same antecedent, $A \Rightarrow D$. Since this is a backwards subjunctive, t_D is before t_A; hence, its default interpretation will put its conditional time some time t^- just before t_D:



Let the function mapping worlds to domains that you get from abstracting on a conditional's domain be its topic: so for a conditional whose domain is $D^{\mathbf{A}}_{w,t,h,g}$, its topic is $\lambda w \,.\, D^{\mathbf{A}}_{w,t,h,g}$. Now, the topics of $A \Rightarrow C$ and $A \Rightarrow D$ in this context are different. The topic of the former is $\lambda w \,.\, D^{\mathbf{A}}_{w,t,h,g}$ and the topic of the latter is $\lambda w \,.\, D^{\mathbf{A}}_{w,t^-,h,g}$. However, it is possible to interpret these conditionals with the same topic—just assign $A \Rightarrow C$ the earlier conditional time. I assume that there is generally pragmatic pressure to interpret subsequent related utterances as about the same topics; this ensures that speakers' contributions will be relevant and coherent.[11] Thus, after accepting an utterance of $A \Rightarrow D$, we will generally interpret $A \Rightarrow C$ as having this earlier conditional time t^-, which will result in a backtracking interpretation.

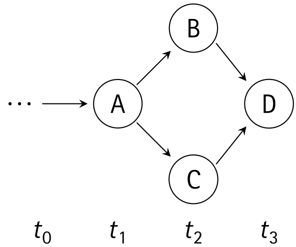
This seems to capture the standard dialogues that involve backtracking interpretations. Recall Lewis's:

> "Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel; *if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday*. In that case Jack would be his usual generous self. **So if Jim asked Jack for help today, Jack would help him after all**." (Lewis 1979a: 456)

Here, the italicized backwards subjunctive shares the same antecedent as the bolded forward subjunctive that ends up receiving the backtracking interpretation.

I should also emphasize that this is not the only way one might make salient the backtracking interpretation of a subjunctive conditional. If the relevant facts about the sufficiency network are common ground, both interpretations seem clearly available. Consider the following case, where each of **A, B, C, D** is true:

[11] I am also open to the possibility that there is a semantic reflex that generates this topic coherence, à la Stojnić 2018, Stojnić et al. 2018.



In this case, the counterfactual:

(22) If B had been false, D would have been false, too.

seems to have a true backtracking reading (as well as a false normal reading). When the sufficiency network is itself common ground, then there is no asymmetry regarding the usefulness of the two kinds of information, and one could be seen as describing one part of the network (the relation between B and D post-t_1) or another (the relation between B and D post-t_0).

Summary: *Whither backtracking*

We argued that, for a metaphysical subjunctive conditional, its conditional time must come before consequent time (in accordance with Avoid Trivialization). This is sufficient to predict the right results for backwards subjunctives. We then saw that backtracking subjunctives result from an earlier than default conditional time, which can be made contextually salient by a preceding backwards subjunctive.

9.6 Elusive Miracles

Similar to the theories of Lewis 1979b, Bennett 2003, assuming deterministic laws of nature, my theory entails that there will be miracles (exceptions to our laws) at the worlds in the domain of a metaphysical subjunctive. That is because my theory does not involve "backtracking unlimited" when evaluating ordinary metaphysical subjunctives (to use Bennett's phrase)—instead, we hold fixed most of the past up until a time very close to when the antecedent is hypothesized to take place, and then generate some break in the sufficiency network to make room for the antecedent. This means that these worlds will often be weird—some may even involve large and unusual miracles, whereby things come into or go out of existence, or suddenly change properties (cf. Jackson 1977). In fact, my embracing of this consequence is slightly different from Lewis and Bennett's approach. Both of them tried to avoid this consequence—Lewis, by appealing to his system of weights, which prioritized small miracles over large ones, and Bennett by demanding some amount of backtracking to find what he calls a "salient admissible fork" allowing for a "smooth" ramp leading from the break in the causal order to the antecedent coming to be.

The kind of case that bothered Bennett was the following (Bennett 2003: 209–210):

> **Dam**. A dam suddenly bursts, quickly submerging a low-lying road and killing the people stuck in their cars in its path.

In this scenario, the following conditional seems true:

(23) If there had been no cars on the road just then, no lives would have been lost in the flood.

But what are the worlds in the domain of (23) like? According to my theory, they are like ours until just before "then" (call this time *t*), at which point the cars stuck in the low-lying road suddenly move somewhere else, resulting in no lives lost. So, I predict (23) is true. Unfortunately, it seems that, for the same reason, I should also predict (24) to be true:

(24) If there had been no cars on the road just then, the cars that were on the road just before would have suddenly moved elsewhere.

After all, I just conceded that this happens at the worlds in the domain of (23). Bennett rightly recognizes this prediction to be problematic, and similar observations have been made by Lange 2000, Dorr 2016. Instead, something more like (25) seems correct:

(25) If there had been no cars on the road just then, the cars that were there would have had some other reason to be somewhere else.

By allowing for a slightly earlier conditional time, we could predict a similar result. However, taking this route seems to open a backdoor for backtracking unlimited. For there were reasons for the reasons the cars had to be somewhere else, etc. For instance, maybe Sue was only on that road at that time because she was using her mapping software, which took her on a special route she normally does not take. Then, had Sue not been on that road, she would not have been using her mapping software. But why was she not using it? We are quickly hurtling towards backtracking unlimited:

(26) If there had been no cars on the road just then, the history of the world would have been different.

One response is to accept this pill and try to finesse this result—perhaps the history of the world would have been different, but not that much different (cf. Dorr 2016). Bennett thinks we can find a principled difference that distinguishes (25) from the line of reasoning that led us to (26), but in the end is unable to say more than that the conditional time is some late salient admissible one.[12]

Instead, I propose that the default conditional time for (23) determines a domain, all of whose worlds are such that the cars on the road suddenly move elsewhere just before its antecedent. However, I also propose that (24) is false and instead (25) is true. How can I do this? Here, in general, is the reason. Let *S* be some statement of some

[12] In Khoo 2017, I defended a version of this Bennettian response, but I am no longer convinced of this strategy for the following reason. Basically, I proposed that conditional time is the latest admissible time (that is, before t_A and t_C), which is such that any later admissible time makes no difference to the conditional's truth value (cf. world w_1 and w_2 above). But this suggests that an interpretive process is sensitive to the actual truth value of the conditional, which depends on the sufficiency structure of the world, and this, I think, is implausible.

true sufficiency, and consider the conditional $A \Rightarrow \neg S$. By (CT1), its conditional time t must be before any proposition involved in S. Then, the conditional's domain $D^{\mathbf{A}}_{w,t,h,g}$ will entail the sufficiency described by S, since it is about times after t, and hence $A \Rightarrow \neg S$ will be false.

Since this holds for any true sufficiency, it explains why (24) is false. "The cars that were on the road then suddenly moved elsewhere" describes the negation of the sufficiency $[\mathbf{C}^{\mathbf{t_1}}] \mapsto \mathbf{C}^{\mathbf{t_2}}$ (basically, if the cars are on the road at t_1, then they are on the road at the next moment t_2). By (CT1), the conditional time of (24) must be before t_1—say, it is t_0. Then $[\mathbf{C}^{\mathbf{t_1}}] \mapsto \mathbf{C}^{\mathbf{t_2}}$ will concern times entirely after conditional time, and thus must be in $h^{\bar{\mathbf{S}}}_{w,t_0,h,g}$, in which case the conditional's domain will entail this sufficiency, and hence it will be false.[13]

Summary: *Elusive miracles*

We saw how our theory predicts that the miracles posited in the worlds making up the domain of a subjunctive conditional will be elusive—we know they are there (when the laws are deterministic), but when we try to state that they are there using a subjunctive conditional $A \Rightarrow \neg S$, where S is some sufficiency, the resulting subjunctive will be false because its conditional time must come before S and thus S will be held fixed at the conditional's domain.

[13] What about conditionals that state what would be consequences of the miracle? For instance:

(i) If there had been no cars on the road just then, there would be more believers in God.

The idea here is that if the cars all miraculously relocated just before the dam burst, more people would believe in God, thinking this an act of God. But for this conditional to be true, it would require there to be a sufficiency like the following, $\begin{bmatrix} \mathbf{C}^{\mathbf{t_1}} \\ \mathbf{C}^{\bar{\mathbf{t}}_2} \end{bmatrix} \mapsto \mathbf{G}$. But given the sufficiency $[\mathbf{C}^{\mathbf{t_1}}] \mapsto \mathbf{C}^{\mathbf{t_2}}$, any conditional time $t : t_1 < t < t_2$ needed to make the antecedent of $\begin{bmatrix} \mathbf{C}^{\mathbf{t_1}} \\ \mathbf{C}^{\bar{\mathbf{t}}_2} \end{bmatrix} \mapsto \mathbf{G}$ true will force us to suspend this sufficiency, since the latter will not be a post-t sufficiency. Thus, we predict conditionals that state what would be consequences of the miracle will also be predicted not to be determinately true. Thanks to Paolo Santorio for discussion on this point.

9.7 Beyond the Possible

I want to close with a speculative extension of our theory to counterpossible subjunctives, which are a well-known fly in the ointment of standard possible-worlds theories of conditionals (see Cohen 1988, Nolan 1997, Goodman 2004, Krakauer 2012, Bjerring 2013, Brogaard & Salerno 2013, Jenny 2018). My suggestion is that the sufficiency structure can be extended into a multi-dimensional framework that will allow us to capture certain similarities between ordinary and counterpossible subjunctives. Start with a couple of plausibly non-vacuous counterpossibles (here, I assume that sets exist necessarily and that origin essentialism and the law of non-contradiction are necessarily true):

(27) If there had been no sets, Obama would still have existed and there would have been no set {Obama}.

(28) If Obama had been born of different parents, he would still have been president.

(29) If Obama had been tall and not tall, Graham Priest would have been correct.

None of these seem vacuously true. Of course, it may be that the source of such intuitions is pragmatic, rather than semantic (see, for instance, Williamson 2015). However, consider the following observation: just as with non-counterpossible subjunctives, there seems to be a contrast between forward and backwards counterpossibles. Suppose Smith thinks that 2 + 2 = 5. Then, we might agree that:

(30) If two plus two had equaled five, Smith would have been correct.

But now flip antecedent and consequent:

(31) a. If Smith had been correct about what two plus two equals, two plus two would still not have equaled five.
b. If Smith had been correct about what two plus two equals, two plus two would have equaled five.

Notice that (31-b) is a non-counterpossible (since neither its antecedent nor consequent are impossible), whereas (31-a) is a counterpossible since its consequent is impossible. Yet, it seems to me that neither seems clearly true: If Smith had been correct, would he have thought that 2 + 2 = 4 or would he have thought what he did and 2 + 2 equaled 5? Neither answer seems clearly true.

Here is another example that exhibits the same behavior:

(32) If there had been a violation of the Law of Excluded Middle, there would have been a violation of a logical law.

(33) a. If there had been a violation of a logical law, there would have been a violation of the Law of Excluded Middle.
b. If there had been a violation of a logical law, there would have been a violation of the Law of Non-Contradiction.

Here, (32) seems true, but neither (33-a) nor (33-b) seem clearly true.

These cases are analogous to David Lewis's barometer example (Lewis 1973a, see also Kratzer 2012: 149). Suppose the current air pressure is 1005.2 millibars and we have a functioning barometer that currently reads 1005.2 millibars. Consider:

(34) If the air pressure had been higher, the barometer reading would have been higher.

This seems true. But now consider its backwards counterpart:

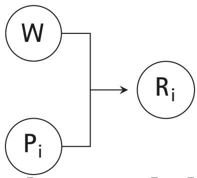
(35) If the barometer reading had been higher, the air pressure would have been higher.

By contrast, (35) seems not clearly true; indeed, it has no more claim to truth than:

(36) If the barometer reading had been higher, the barometer would have been malfunctioning.

The contrast between (30) and (31-a) is unexpected if we thought that counterpossibles were merely a pragmatic phenomenon, and it is not easy to explain given the natural thought that impossible worlds are automatically ranked as less similar than possible worlds (though see Nolan 1997: 550–551, and Vander Laan 2004: 271).

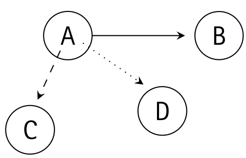
Instead, appealing to sufficiency networks seems like a plausible account of this contrast. Consider the barometer case, where the background facts plausibly determine something like the following sufficiency network (here, **W** = the barometer is working, P_i = the air pressure is i, and R_i = the barometer reads i):



$$\mathcal{S} = \{\begin{bmatrix}\mathbf{W}\\ \mathbf{P_1}\end{bmatrix} \mapsto \mathbf{R_1}, \ldots, \begin{bmatrix}\mathbf{W}\\ \mathbf{P_n}\end{bmatrix} \mapsto \mathbf{R_n}\}$$

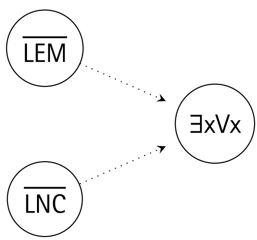
In this case, the domain $D^{\mathbf{P}_{+1}}_{w,t,m,g}$ will entail all of the sufficiencies in $\mathcal{S}$ and thus entail R_{+1}. And similarly the domain $D^{\mathbf{R}_{+1}}_{w,t,m,g}$ will entail all of the sufficiencies in $\mathcal{S}$, but won't entail P_{+1} or **W**, but only their disjunction.

To account for this parallel, we allow that, just as propositions are related via causal sufficiency relations (ordered according temporal order), they are also related via grounding sufficiency relations (ordered via priority; see Schaffer 2009, 2016; Sider 2013; Baron et al. 2017). We can then intervene in the priority sufficiency network, just as we intervene in the causal sufficiency network. And, indeed, we may in some cases need to intervene in both simultaneously. Thus, imagine a multi-dimensional network of propositions—here is a simple example involving three dimensions of sufficiencies:



Here, **A** is sufficient for **B**, **C**, and **D**, but in different dimensions: perhaps causal, grounding, and explanatory sufficiency. That means that, although **A** is in some way sufficiency-related to each of these propositions, they are not sufficiency-related to each other (different kinds of sufficiency block transitivity and, in this case, Euclideanness). It also means that interventions in one dimension are independent of the other dimensions. Notice as well that times keep track both of temporal order (as in the causal sufficiency dimension), as well as other kinds of priority. So, if **A** is before **B** in the grounding network, this means that **A** is grounding prior to **B**, even though **A** may not be temporally prior to **B**. This means also that times are now really multi-dimensional points—tuples of times in each sense of order. So, where before we had t_1, which was a time in the causal temporal network, we now have $\langle t_1, t_a, t_i \rangle$, where t_a and t_i are nodes in the grounding and explanatory networks, respectively (etc.).

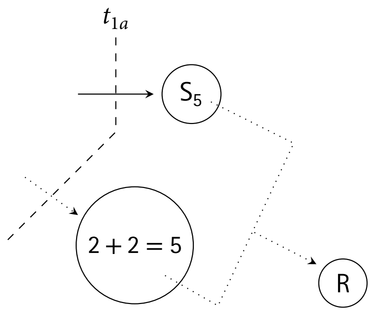
Let us look at (32) as an example. This conditional involves interventions just in the grounding network. We suppose the relevant part of the grounding network looks like this (representing that a violation in the Law of Excluded Middle or the Law of Non-Contradiction is grounding sufficient for there being some violation of a logical law):



Given this network, intervening just before $\overline{\textbf{LEM}}$ to set it to true also requires setting $\exists x Vx$ to true, which is how we predict the truth of (32). But (33-a)/(33-b) are different. Intervening before both and setting $\exists x Vx$ to true require setting one of them to true, but does not specify which. Thus, we predict that neither (33-a) nor (33-b) are factually true.

Now turn to (30). In this case, our sufficiency network crosscuts two dimensions ($\mathbf{S_5}$ = Smith thinks 2 + 2 = 5, **R** = Smith was right;

note that just as we have the sufficiency $\begin{bmatrix}\mathbf{S_5}\\ \mathbf{2+2=5}\end{bmatrix}\mapsto \mathbf{R}$, we also have the unillustrated $\begin{bmatrix}\mathbf{S_4}\\ \mathbf{2+2=4}\end{bmatrix}\mapsto \mathbf{R}$):



Intervening at t_{1a} to set $\mathbf{2 + 2 = 5}$ to true ensures that **R** will also be true, thus predicting (30). But now notice that we can predict that both (31-a) and (31-b) are non-factual. Intervening at t_{1a} to set **R** to true demands that either $\mathbf{2 + 2 = 5}$ be true, or $\mathbf{S_4}$ be true, yet neither change is necessitated. Interestingly, if we intervened at t_{1b}, which is some point in the network before $\mathbf{S_5}$ is settled but at which $\mathbf{2 + 2 = 4}$ is settled, the conditional (37) would be true:

(37) If Smith had been correct, he would have thought 2 + 2 = 4.

That is a welcome prediction—and notice that we do predict this, because, unlike with (31-a)/(31-b), there is no demand that the conditional time of (37) be before $\mathbf{2 + 2 = 4}$ is settled.

It remains to be seen whether this extension of the theory developed in §8.3 will withstand further scrutiny, but for now I hope to have at least motivated the view that it is worth taking seriously among various options for handling counterpossibles.[14]

[14] Notice, furthermore, that the resulting theory allows that tense can shift dimensions of priority that do not correspond to event times, but this is still not a modal theory of tense. Rather, it might be best thought of a multi-dimensional theory of tense.

9.8 Summary

This chapter has focused on the meanings of metaphysical subjunctive conditionals. I argued that such conditionals express information about sufficiency networks (§9.1). However, the standard theory connecting conditionals and sufficiencies—Interventionism—faces challenges stemming from backwards and backtracking subjunctives (§9.2–9.3). By contrast, the temporal past theory from Chapter 8 provided a strategy allowing us to divorce the intervention point from the conditional's antecedent and consequent—the resulting theory inherits the benefits of Interventionism without its flaws (§9.4). I then offered explanations of the pragmatics of conditional time that explained why backtracking interpretations are non-default and when they arise (§9.5), and why miracles are counterfactually elusive, even though my theory predicts almost all subjunctive domains entail them (§9.6). I concluded with a discussion of how to extend this theory to counterpossibles, arguing that it could predict similar contrasts for both non-counterpossible and counterpossible forward and backwards subjunctives (§9.7).

One outstanding issue that has loomed large in our discussion is the fact that not all subjunctive conditionals are factually true—indeed, some have domains that do not settle their consequents to be true or to be false. In that case, what is the status of the subjunctive? Philosophers like Lewis thought that a subjunctive with this property was false. But that is not the theme of this book. Instead, as with indicatives, I hold that such conditionals are non-factual, and that this non-factuality non-trivally affects their probabilities. In our final chapter, I will bring together my theory of subjunctive conditionals with my discussion of indicative conditionals in Chapters 3–6, showing how we can predict non-trivial subjunctive probabilities.

10
Subjunctive Probabilities

Like indicative conditionals, subjunctive conditionals often have non-extremal probabilities.

> **Case 1**. Smith's coin was either double-headed or double-tailed, and we think each possibility is 50% likely. We know Smith did not flip the coin.

On this basis, it seems we should think that both of the following are 50% likely:

(1) a. If Smith had flipped the coin, it would have landed heads.
    b. If Smith had flipped the coin, it would have landed tails.

After all, if Smith's coin is double-headed, then (1-a) is factually true, and if Smith's coin is double-tailed, then (1-b) is factually true, and those possibilities are equally likely and exhaust the possibilities compatible with our evidence.

Now consider a slight variation on the case:

> **Case 2**. We know that Smith holds a chancy fair coin, one that is such that flipping it does not deterministically cause it to land heads nor deterministically cause it to land tails, though flipping it does cause it to land one way or the other. We know that the chance of this coin landing heads if flipped is equal to the chance of it landing tails if flipped, and that those are the only two possible outcomes. Again, we know that Smith did not flip the coin.

On this basis, it seems we should again think that both of (1-a)/(1-b) are 50% likely. However, in this case, the uncertainty of these conditionals does not stem from our ignorance of some matter of fact that

The Meaning of If. Justin Khoo, Oxford University Press. © Oxford University Press 2022.
DOI: 10.1093/oso/9780190096700.003.0011

entails the factual truth of one of them, but rather our knowledge of the chances that each is true.[1]

Finally, there are also mixed cases:

Case 3. It is 50% likely that Smith has a chancy fair coin, and 50% likely that he has a double-headed coin.

Then, it seems that we should think that (1-a) is $3/4$ likely, while (1-b) is $1/4$ likely. The intuitive calculation is clear: it is $1/2$ likely that Smith has a double-headed coin, in which case (1-a) would be definitely true; and it is $1/2$ likely that Smith has a chancy fair coin, in which case the chance of (1-a) would be $1/2$, meaning that its overall probability should be $3/4$.[2]

Ideally, our theory would account for both sources of uncertainty, as well as their interaction. But there are problems: just like with epistemic indicatives, triviality results threaten the possibility of non-trivial probabilities for metaphysical subjunctives. In this chapter, I show how to adapt the results from the previous chapters to metaphysical subjunctives, and prove a tenability result that generates what I will argue to be plausible probabilities for such conditionals. In §9.1, I draw out the generalizations about the probabilities of subjunctives that we will aim to predict. Then, in §9.2, I discuss two triviality results for subjunctive conditionals (one from Williams 2012 and one from Santorio 2021), which put pressure on our observations. I draw on the strategy of Moss 2013 to defend against Williams's result, and then show our already motivated rejection of Closure under Conditionalization helps to defend us against Santorio's result. Finally, in §9.3, I prove a tenability result for subjunctives (extending the result from Chapter 4) that shows my theory predicts these observations.

10.1 Probabilities for Subjunctives

The above observations suggest that something like Skyrms's Thesis is the correct observation for the probabilities of subjunctive

[1] Note that chancy conditionals can arise even in a deterministic setting. We saw how with backwards subjunctives in Chapter 9.

[2] I should point out that I will be taking these judgments at face value. For an argument in their favor, see Moss 2013; for an argument against, see Hájek 2020.

conditionals. Here, Ch_w is the objective chance function of the world w, and $[Ch(\mathbf{A}) = x]$ is a shorthand for $\{w : Ch_w(\mathbf{A}) = x\}$.[3]

Skyrms's Thesis
$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = \sum_x Pr([Ch(\mathbf{B}|\mathbf{A}) = x]) \cdot x$$

Moss 2013 argues for a restricted version of Skyrms's Thesis as applied to future-oriented subjunctives, and even shows how to derive Skyrms's Thesis, which she calls Bridge, from the Principle Principal. Reflection on the cases above suggests that we should endorse Skyrms's Thesis generally (for both past and future subjunctives).

In Case 1, there are two possible chance-hypotheses. At worlds w where the coin is double-headed, $Ch_w(\mathbf{H}|\mathbf{F}) = 1$; and at worlds w' where the coin is double-tailed, $Ch_{w'}(\mathbf{H}|\mathbf{F}) = 0$. Since each chance-hypothesis is equally likely, Skyrms's Thesis predicts that $Pr(\mathtt{F} \Rightarrow \mathtt{H}) = 1/2$.

In Case 2, there is one possible chance-hypothesis: since we know the coin is fair, $Ch_w(\mathbf{H}|\mathbf{F}) = 1/2$. Since this chance-hypothesis is believed with certainty, Skyrms's Thesis predicts that $Pr(\mathtt{F} \Rightarrow \mathtt{H}) = 1/2$.

Finally, in Case 3, there are two possible chance-hypotheses: at w where the coin is double-headed, $Ch_w(\mathbf{H}|\mathbf{F}) = 1$; and at w' where the coin is fair, $Ch_{w'}(\mathbf{H}|\mathbf{F}) = 1/2$. Since each is equally likely, Skyrms's Thesis predicts that $Pr(\mathtt{F} \Rightarrow \mathtt{H}) = 3/4$.

However, despite the success here of Skyrms's Thesis, it cannot hold in full generality. Like Moss, I am convinced that it can hold at most for future-oriented subjunctives. The problem comes from Morgenbesser-style cases (discussed extensively in Chapter 9):[4]

Case 4. Jones flips a fair coin, Sue bets tails, the coin lands heads.

Given this information, you know that the pre-flip chance of Sue winning given that she bets tails is 1/2, since the pre-flip chance of

[3] Skyrms's original proposal was about the "assertability value" of subjunctive conditionals, but I will follow Schulz 2017 and work with the formulation in terms of probabilities.

[4] See Edgington 2008, 2014; Schulz 2017 for applications of Morgenbesser cases to challenge Skyrms's Thesis.

the coin landing tails is ½—note you know this even though you know that the coin landed heads (objective chances are not sensitive to inadmissible information—Lewis 1980b). Nonetheless, knowing that the coin did in fact land heads, you should have probability 1 in the conditional:

(2) If Sue had bet tails, she would have won.

Moss points out correctly that this probability is just the result of conditionalizing your earlier probability in:

(3) If Sue were to bet tails, she would win.

whose probability is plausibly constrained by Skyrms's Thesis, on the information that the coin landed heads. Since the coin landing heads entails (2), we thus predict the correct probability for the conditional. I think that is exactly right. But this shows that we actually need to start with the conditional semantics before we get to their probabilities. As Moss shows, starting with the Principle Principal (Lewis 1980b) is sufficient to derive Skyrms's Thesis. But to get a fully general theory of the probabilities of subjunctives, we need to augment this with a theory of their semantics—in particular, a semantic theory that predicts the fact that the coin landed heads entails (2) but not (4):[5]

(4) If Sue had flipped the coin (rather than Jones), she would have won.

In Chapter 9, I defended this kind of semantics, so this puts me in a position to combine this theory with Moss's to generate a fully general theory of the probabilities of subjunctives. But it still remains to be seen whether and how to do so.

One immediate issue is this. In Chapter 5, I showed that, given certain restrictions, the probability of an indicative conditional is equal

[5] An alternative strategy has recently been advocated by Schultheis 2021, which argues that we can derive Skyrms's Thesis from Hall 1994's New Principle plus a local version of Stalnaker's Thesis.

to the conditional probability of the conditional's consequent given its antecedent. Our theory predicted this result by motivating a way of extending our factual probability measure P into a refined probability measure Pr that encoded the probabilistic inferential dispositions generated by P. However, whereas indicative conditionals presuppose that their antecedents are epistemically possible, subjunctives presuppose that their antecedents are metaphysically possible, and many have counterfactual antecedents (antecedents believed or known to be false). But this means that for all worlds w compatible with your factual beliefs, $P(D^{\mathbf{A}}_{w,t,m,g}) = 0$, and the conditional probability $P(\mathbf{B}|D^{\mathbf{A}}_{w,t,m,g})$ will be undefined.

Recall from Chapter 3 that counterfactual inferential dispositions (those dispositions to infer something from a proposition incompatible with your factual beliefs) must not distinguish between possibilities ruled out by your factual beliefs. The motivating example was the case:

> Suppose you think the prize is not behind doors D, E, or F. Then, one is not disposed to infer that the prize is behind door D upon believing it's behind door D or E.

Shifting to probabilities, you also have dispositions to infer various probabilities from propositions that are incompatible with your beliefs. Here, the constraint not to distinguish between possibilities ruled out by your factual beliefs becomes the constraint not to probabilistically distinguish those possibilities. Thus, when $P(\mathbf{A}) = 0$, your counterfactual disposition to infer that **B** has probability x from that **A** should be the probability of randomly selecting a **B**-world from among the **A**-worlds.

Define the ur-probability measure $\mathring{P}$ as one that assigns equal probability to every possible world. We can use this to define the chance of randomly selecting a **B**-world from a set of worlds **A** as follows: $\mathring{P}(\mathbf{B}|\mathbf{A})$. Thus, where P models your credences and $P(\mathbf{A}) = 0$, then you are counterfactually disposed to infer that $P(\mathbf{B}) = x$ from **A** iff $\mathring{P}(\mathbf{B}|\mathbf{A}) = x$. With this device in hand, we can now generalize our theory of indicative probabilities to subjunctives as follows: if

$P(\mathbf{A}) = 0$, then your probability in $\mathtt{A} \Rightarrow \mathtt{B}$ just is your expectation of randomly selecting a **B**-world from the conditional's domain, $D^{\mathbf{A}}_{w,t,m,g}$.[6]

To state this carefully, I will adopt a shorthand to ease the presentation of the theory. So far, we have used the expression '$D^{\mathbf{A}}_{w,t,m,g}$' to refer to the domain of a conditional $A \Rightarrow B$ (whose conditional time is t, modal base m, and ordering source g). Here, I will suppress relativization to conditional time, modal base, and ordering source, and instead use

$$D^{\mathbf{A}}_{w}$$

to refer to the conditional $A \Rightarrow B$'s domain at w. As above, I'll adopt the shorthand of writing:

$$[D^{\mathbf{A}}]$$

to refer to the proposition:

$$\{w : D^{\mathbf{A}}_{w}\}$$

With these simplifications in hand, we can state our principle about the probabilities of subjunctive conditionals as follows:

Subjunctive Probability

For any Pr and non-conditional A, B such that $P(\mathbf{A}) = 0$ and $P([D^{\mathbf{A}} \neq \varnothing]) = 1$,

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = \sum_{x} P([\mathring{P}(\mathbf{B} | D^{\mathbf{A}}_{w}) = x]) \cdot x$$

As it stands, Subjunctive Probability is not fully general, because it doesn't say what happens for non-counterfactual subjunctives, where $Pr(\mathtt{A}) > 0$. I'll return to non-counterfactual subjunctives below. For now, it should be easy to observe that this theory predicts the same

[6] Subjunctive Probability below bears similarities to principles defended by Skyrms 1981, Edgington 2008, Schulz 2017, although I arrive at this result from a very different direction than these authors.

results as Skyrms's Thesis for Cases 1–3. When it comes to the Morgenbesser case, we have the following. At each world in your belief state **BEL**, the coin landed heads. Furthermore, its doing so at those worlds was independent of how Sue bet. Thus, since this is a post-t fact independent of $\mathbf{B_H}$, we hold fixed this fact at the worlds in the conditional's domain: hence, $D_w^{\mathbf{B_H}} \subseteq \mathbf{H}$. And thus we predict that $Pr(\mathtt{B_H} \Rightarrow \mathtt{W}) = 1$.

We are doing well so far. But not all subjunctives are counterfactual. Define:

$$\mathtt{A} \Rightarrow \mathtt{B} \text{ is } \begin{cases} \text{fully counterfactual for } \mathbf{BEL} & \text{if } \mathbf{BEL} \cap \mathbf{A} = \varnothing \\ \text{fully non-counterfactual for } \mathbf{BEL} & \text{if } \mathbf{BEL} \cap \bar{\mathbf{A}} = \varnothing \\ \text{partially counterfactual for } \mathbf{BEL} & \text{otherwise} \end{cases}$$

We have already seen that some subjunctives are partially counterfactual—Stalnaker-type examples like (5) fall into this category:[7]

(5) If the butler had done it, he would have used a knife. (So, let's find out what the murder weapon was.)

To assign probabilities to such subjunctives, we need to take into consideration the fact that, given Strong Centering, a subjunctive $A \Rightarrow B$ is factually true when **AB** is true, and factually false when $\mathbf{A\bar{B}}$ is true. We also need to consider the relationship between the subjunctive's domain and the agent's belief state at worlds in the belief state where its

[7] Larry Horn (personal communication) pointed me to a fully non-counterfactual subjunctive from Diana Gabaldon's 2004 novel *The Fiery Cross*:

> "Any chance your uncle Hector was an opium-eater or the like?[. . .]"
>
> Never one to be shocked by any intimation of depravity among his relatives, Jamie considered the suggestion but finally shook his head. "If so, I've heard naught of it. But then," he added logically, "there's no reason anyone would tell me so."
>
> That was true enough. **If Hector Cameron had had the means to indulge in imported narcotics—and he certainly had, River Run being one of the most prosperous plantations in the area—then it would have been nobody's business but his own.** (Gabaldon 2005: 1240)

I will set aside examples like this, which are predicted to be as probable as their consequents.

antecedent is false. If the domain is compatible with the belief state at some such worlds, then only the belief-worlds in the domain matter to the probability of the subjunctive. But if the subjunctive is orthogonal to the belief state at such worlds, then every world in the domain matters to the probability of the subjunctive. In what follows, I will suppose that a subjunctive domain will be orthogonal to the agent's belief state at worlds where its antecedent is false. This is plausible for past subjunctives, as can be seen by the following example:

> You know the coin was fair, and also that it is 1/2 likely that it was flipped and landed tails and 1/2 likely that it was not flipped at all.

How probable should you regard:

(6) If the coin was flipped, it landed heads.

(7) If the coin had been flipped, it would have landed heads.

Start with the indicative (6). Since the only open possibilities in which the coin was flipped, it landed tails, you should regard (6) as certainly false. But things are different with the subjunctive (7). It is false at worlds in which the coin was flipped and landed tails, but at the worlds where it wasn't flipped, since it was fair, it should have a probability of 1/2; therefore, its overall probability should be 1/4. However, this reasoning only holds if all of the worlds in the subjunctive's domain (at worlds where it wasn't flipped) matter—in which case, its domain must be orthogonal to our belief state at such worlds.

Given orthogonality, the general equation is thus:

Generalized Subjunctive Probability

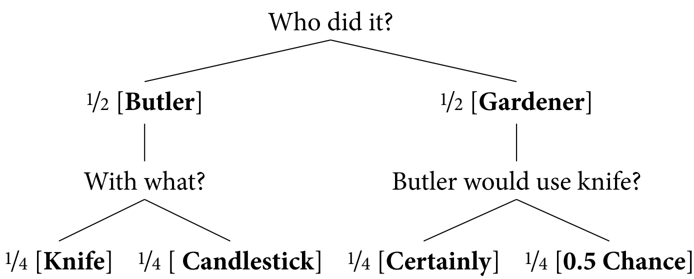
For any Pr and non-conditional A, B such that $P([D^{\mathbf{A}} \neq \varnothing]) = 1$,

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = P(\mathbf{AB}) + \Big[\sum_x P([\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = x] \mid \bar{\mathbf{A}}) \cdot x\Big] \cdot P(\bar{\mathbf{A}})$$

Turn back to (5):

(5) If the butler had done it, he would have used a knife. (So, let us find out what the murder weapon was.)

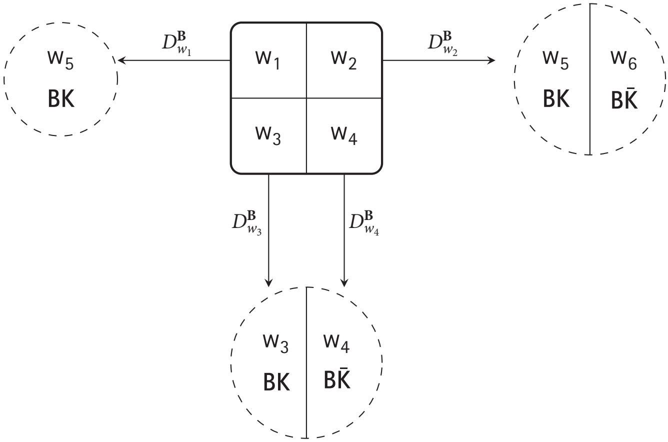
Suppose there are two suspects, the butler and the gardener (each 1/2 likely to have done it), and that you think it is 1/4 likely that the butler did it with a knife and 1/4 likely he did it with a candlestick. Furthermore, you also think the butler and gardener were in cahoots so that if the gardener did it, the butler was an accomplice, ready to jump in as a backup. You think it is 1/2 likely that, given the gardener did it, he arranged for the butler to use a knife as a backup weapon, and 1/2 likely that, given the gardener did it, he did not arrange for the butler to use a specific backup weapon, in which case the butler flipped a fair coin and was prepared to use a knife on heads and a rope on tails. Thus:



On this basis, it seems that the probability of (5) is around 5/8. Why? Well, we have the 1/4 probability that the butler used the knife, as well as the 3/8 probability that the gardener did it and the butler would have used a knife as a backup (this comes from the certain chance he would have used a knife as backup = 1/4 plus the 0.5 chance he would have used the knife as backup = 1/8).

Given Generalized Subjunctive Probability, we predict this result:

- **Case 5 (Partially Counterfactual, Mixed Factual/Non-Factual)**
  - $\mathbf{BEL} = \{w_1, w_2, w_3, w_4\}$
  - $D^{\mathbf{B}}_{w_1} = \{w_5\}$, $D^{\mathbf{B}}_{w_2} = \{w_5, w_6\}$, $D^{\mathbf{B}}_{w_3} = \{w_3, w_4\}$, $D^{\mathbf{B}}_{w_4} = \{w_3, w_4\}$
  - $\mathbf{B} = \{w_3, w_4, w_5, w_6\}, \mathbf{K} = \{w_3, w_5\}$



$$Pr(\texttt{B} \Rightarrow \texttt{K}) = \underbrace{P(\mathbf{BK})}_{1/4} + \Big[\sum_x P([\mathring{P}(\mathbf{K}|D^{\mathbf{B}}_w) = x] \mid \bar{\mathbf{B}}) \cdot x\Big] \cdot \underbrace{P(\bar{\mathbf{B}})}_{1/2}$$

$$\mathring{P}(\mathbf{K}|D^{\mathbf{B}}_{w_1}) = 1$$

$$\mathring{P}(\mathbf{K}|D^{\mathbf{B}}_{w_2}) = 1/2$$

Hence,

$$P([\mathring{P}(\mathbf{K}|D^{\mathbf{B}}_w) = 1] \mid \bar{\mathbf{B}}) = 1/2$$

$$P([\mathring{P}(\mathbf{K}|D^{\mathbf{B}}_w) = 1/2] \mid \bar{\mathbf{B}}) = 1/2$$

So,

$$Pr(\texttt{B} \Rightarrow \texttt{K}) = 1/4 + \underbrace{(1 \cdot 1/2 + 1/2 \cdot 1/2)}_{3/4} \cdot 1/2 = 5/8$$

In what follows, I will set aside these non-counterfactual subjunctives to simplify our calculations. However, despite the plausible predictions made by our theory so far, a serious problem looms—a threat from triviality for subjunctive conditionals to which we turn next.

10.2 Triviality for Subjunctives

Before I show how we can predict Subjunctive Probability, in this section I want to diffuse two triviality results which threaten to establish that it cannot be correct.

10.2.1 Williams 2012

Williams 2012 argues that Lewis's original 1976 diachronic triviality result can be extended to subjunctive conditionals. His strategy is to lift the proof to chance distributions, using Lewis 1980b's Principal Principle and the assumption that the relevant chances are known:

Principal Principle

Where P is the probability function of an agent with no inadmissible evidence,

$$P(\mathbf{A}) = \sum_{x} P([Ch(\mathbf{A}) = x]) \cdot x$$

The basic idea here is that agents set their credences to their expectation of the objective chances (assuming they lack inadmissible information, such as information about the future from a reliable oracle).

Suppose that you know the objective chance of $\mathtt{A} \Rightarrow \mathtt{B}$:

$$P(Ch(\mathtt{A} \Rightarrow \mathtt{B}) = z) = 1$$

It follows from the Principal Principle that:

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = z$$

Furthermore, suppose that:

$$P(\mathbf{A}) = 0$$

Then, it follows from Subjunctive Probability above that:

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = \sum_x P([\mathring{P}(\mathbf{B}|D_w^{\mathbf{A}}) = x]) \cdot x = z$$

But let's also suppose that you know what the domain of the conditional is:

$$P([\mathring{P}(\mathbf{B}|D_w^{\mathbf{A}}) = x]) = 1$$

It follows that:

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = \mathring{P}(\mathbf{B}|D_w^{\mathbf{A}}) = z$$

and thus that:

$$Ch(\mathtt{A} \Rightarrow \mathtt{B}) = \mathring{P}(\mathbf{B}|D_w^{\mathbf{A}})$$

Now, we are in business. We can appeal to analogues of Lewis's Lemmas:

(S1) $Ch(\mathtt{A} \Rightarrow \mathtt{B}|\mathtt{B}) = \mathring{P}(\mathbf{B}|D_w^{\mathbf{A}} \cap \mathbf{B}) = 1$

(S2) $Ch(\mathtt{A} \Rightarrow \mathtt{B}|\bar{\mathtt{B}}) = \mathring{P}(\mathbf{B}|D_w^{\mathbf{A}} \cap \bar{\mathbf{B}}) = 0$

and then compute via Total Probability:

$$Ch(\mathtt{A} \Rightarrow \mathtt{B}) = \underbrace{Ch(\mathtt{A} \Rightarrow \mathtt{B}|\mathtt{B})}_{1} \cdot Ch(\mathtt{B}) + \underbrace{Ch(\mathtt{A} \Rightarrow \mathtt{B}|\bar{\mathtt{B}})}_{0} \cdot Ch(\bar{\mathtt{B}}) = Ch(\mathtt{B})$$

in which case it follows that:

$$P(Ch(\mathtt{A} \Rightarrow \mathtt{B}) = Ch(\mathtt{B})) = 1$$

and thus from two instances of the Principal Principle that:

$$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = Pr(\mathtt{B})$$

This result is clearly wrong. We know we didn't flip the coin, and thus that it didn't land heads and didn't land tails. So the probability that the coin landed heads is 0. Nonetheless, since we know that the coin is fair, the probability that if we had flipped the coin, it would have landed heads is $1/2$.

Following Moss 2013, I think we ought to reject the two lemmas (S1) and (S2) (a similar point is made by Schulz 2017: 208–209). Moss provides a useful counterexample on pp. 273–274 that I adapt here:

> Smith flips a fair coin, but it hasn't landed yet. The chance that it lands heads is $1/2$ and the chance of tails is $1/2$. Suppose it in fact lands heads. Then, the chance that if Jones (rather than Smith) had flipped the coin, it would have landed heads is still $1/2$. (This same thing holds if it in fact lands tails.)

What I find so interesting about this observation of Moss's is that the analogous story for indicative conditionals and epistemic probability does not work this way:

> Someone flips a fair coin but we don't know the outcome yet. With no evidence, we think it is $1/2$ likely to be heads and $1/2$ likely to be tails. Suppose it in fact landed heads. In that case, there is some intuitive pull to think that the probability that if Jones flipped the coin, it landed heads, should be 1. (The same thing holds if it in fact lands tails.)

In the indicative case, we need to account for the fact that cases like these seem to motivate thinking that (L1) is valid:

(L1) $Pr(\mathtt{A} \Rightarrow \mathtt{B} | \mathtt{B}) = P(\mathbf{B} | \mathbf{AB}) = 1$

My strategy accounts for the appeal of (L1) by suggesting that we often fail to distinguish between conditionalization and learning, which I argued (following the rejection of Conditionalization^{-}) can come apart. My point here is that the principles (S1)/(S2) are by contrast much less plausible to begin with. Thus, a theory that predicts (S1) and (S2) are invalid (thus avoiding Williams's adaptation of Lewis's triviality result) seems to be in good intuitive shape as well.

Now, one might ask whether the theory sketched here is compatible with the Principal Principle, in light of the rejection of (S1)/(S2). And here again I will appeal to Moss 2013 and her derivation of Skyrms's Thesis for future-directed subjunctives from the Principal Principle (see her derivation of BRIDGE on p. 262). Moss's derivation appeals to an independence assumption—that the chance of $A \Rightarrow B$ is equal to the chance of $A \Rightarrow B$ given A. And she shows convincingly that the conditionalized chance functions in (S1)/(S2) involve failures of independence, thus blocking her derivation from entailing these principles.

Thus, drawing on the strategy motivated by Moss 2013, we can avoid the triviality result coming from Williams 2012.

10.2.2 Santorio's Triviality Result

Recently, Paolo Santorio has offered a new triviality result for metaphysical subjunctives that does not rely on detouring through chances (Santorio 2021). Here, I simplify Santorio's reasoning to establish the following (as before, $A \mathrel{\Box\!\!\rightarrow} B$ is the strong conditional corresponding to $A \Rightarrow B$; so $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B} = \{w : D_w^{\mathbf{A}} \subseteq \mathbf{B}\}$):[8]

> **Santorio Triviality**
> For all A, B and Pr that encodes the probabilistic inferential dispositions of a possible agent, if $Pr(\mathtt{A} \mathrel{\Box\!\!\rightarrow} \mathtt{B}) = 0$, then $Pr(\mathtt{A} \Rightarrow \mathtt{B}) = 0$.

[8] Santorio's result establishes that $Pr(\mathtt{A} \Rightarrow \mathtt{B}) = P(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B})$. The full result requires the following additional assumptions:

(i) $P(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \bar{\mathbf{B}}) > 0$

(ii) $P(\neg(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \bar{\mathbf{B}})) > 0$

(iii) **Upper Bound**
For all A, B and Pr: If $P(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}) = 1$, then $Pr(\mathtt{A} \Rightarrow \mathtt{B}) = 1$

See Santorio 2021 for details.

That this is already problematic should be clear. Jones never flipped the fair coin, but it seems clear that (8) has probability 1/2. Yet, since the coin was fair, its antecedent doesn't entail its consequent, so the corresponding strong conditional must have probability 0.

(8) If Jones had flipped the coin, it would have landed heads.

As a shorthand, let $A \diamondright B$ be the sentence schema whose instances have as their content: $\neg(\mathbf{A} \boxright \bar{\mathbf{B}})$. The assumptions we need to derive Santorio Triviality are:

Non-Zero

For all A, B and all Pr encoding the probabilistic inferential dispositions of a possible agent, such that $Pr(\mathtt{A} \diamondright \mathtt{B}) > 0$:

$$Pr(\mathtt{A} \Rightarrow \mathtt{B} | \mathtt{A} \diamondright \mathtt{B}) > 0$$

Conditional Non-Contradiction (CNC)

$\mathtt{A} \Rightarrow \bar{\mathtt{B}} \vDash \neg(\mathtt{A} \Rightarrow \mathtt{B})$

Closure Under Conditionalization

If Pr encodes the probabilistic inferential dispositions of a possible agent and $Pr(\mathtt{A}) > 0$, then $Pr(\cdot|\mathtt{A})$ encodes the probabilistic inferential dispositions of a possible agent.

Let Pr encode the probabilistic inferential dispositions of an agent, and suppose $Pr(\mathtt{A} \boxright \mathtt{B}) = 0$, and thus $Pr(\mathtt{A} \diamondright \bar{\mathtt{B}}) = 1$.

1. $Pr(\mathtt{A} \Rightarrow \mathtt{B}) > 0$ — Reductio assumption
2. $Pr(\mathtt{A} \Rightarrow \mathtt{B} \wedge \mathtt{A} \diamondright \bar{\mathtt{B}}) > 0$ — Setup, 1
3. $Pr(\cdot|\mathtt{A} \Rightarrow \mathtt{B})$ encodes the probabilistic inferential dispositions of a possible agent. — Closure, Setup, 2
4. $Pr(\mathtt{A} \diamondright \bar{\mathtt{B}}|\mathtt{A} \Rightarrow \mathtt{B}) = 1$ — Setup, 3
5. $Pr(\mathtt{A} \Rightarrow \bar{\mathtt{B}}|\mathtt{A} \Rightarrow \mathtt{B}) > 0$ — Nonzero, 3, 4
6. $Pr(\mathtt{A} \Rightarrow \bar{\mathtt{B}}|\mathtt{A} \Rightarrow \mathtt{B}) = 0$ — CNC

Santorio's result lends further support to our theory, which motivates principled counterexamples to Closure Under Conditionalization (see also the discussion in Chapters 4 and 5) and thus blocks the result at Step 3. Here is a case to illustrate.

Take the Jones coin flip case from above, where we know the coin was fair and unflipped. On this basis, we fully believe $\{w : D_w^{\mathbf{F}} \not\subseteq \mathbf{H}\}$ and thus, $Pr(\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}}) = 1$. Furthermore, $Pr(\texttt{F} \Rightarrow \texttt{H}) = 1/2$, since we expect to be disposed to infer $P(\mathbf{H}) = 1/2$ from $D^{\mathbf{F}}$. We now wonder whether $Pr(\cdot|\texttt{F} \Rightarrow \texttt{H})$ encodes the probabilistic inferential dispositions of any possible agent. I will argue that it does not.

Since $Pr(\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}}) = 1$, our agent cannot come to accept $\texttt{F} \Rightarrow \texttt{H}$. Why? Remember, to accept $\texttt{F} \Rightarrow \texttt{H}$ is to be disposed to infer $\mathbf{H}$ from $D_w^{\mathbf{F}}$ for each doxastically possible w. But at every doxastically possible w, $D_w^{\mathbf{F}} \not\subseteq \mathbf{H}$. And, furthermore, since your inferential dispositions cannot distinguish between worlds you've ruled out, you can be disposed to infer $\mathbf{H}$ from $D_w^{\mathbf{F}}$ only if $D_w^{\mathbf{F}} \subseteq \mathbf{H}$. But since that holds at none of your doxastically possible worlds, you cannot come to accept $\texttt{F} \Rightarrow \texttt{H}$ (without revising your belief in $\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}}$). In other words, there is no function Pr modeling the probabilistic inferential dispositions of an agent where $Pr(\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}}) = 1$ and $Pr(\texttt{F} \Rightarrow \texttt{H}) = 1$. Yet, $Pr(\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}}|\texttt{F} \Rightarrow \texttt{H}) = 1$ and $Pr(\texttt{F} \Rightarrow \texttt{H}|\texttt{F} \Rightarrow \texttt{H}) = 1$. Therefore, $Pr(\cdot|\texttt{F} \Rightarrow \texttt{H})$ cannot model the probabilistic inferential dispositions of any possible agent.

It may help to see things model theoretically. Suppose:

$$\begin{array}{ll} \mathbf{W} = \{w_1, w_2, w_3\} & D_{w_1}^{\mathbf{F}} = \{w_2, w_3\} \\ \mathbf{BEL} = \{w_1\} & \mathbf{H} = \{w_2\} \\ & \mathbf{T} = \{w_3\} \end{array}$$

Your refined belief state is:

$$\Uparrow\mathbf{BEL} = \left\{ \begin{array}{c} \langle w_1, w_2, w_3 \rangle \\ \langle w_1, w_3, w_2 \rangle \end{array} \right\}$$

We have:

$$\texttt{F} \diamond\!\!\rightarrow \bar{\texttt{H}} = \left\{ \begin{array}{c} \langle w_1, w_2, w_3 \rangle \\ \langle w_1, w_3, w_2 \rangle \end{array} \right\}$$

$$\mathtt{F} \Rightarrow \mathtt{H} = \{\langle w_1, w_3, w_2\rangle\}$$

Our agent's probability measure *Pr* assigns each sequence in ⤊**BEL** equal probability = 1/2. This, again, encodes the fact that agent's inferential dispositions must not probabilistically distinguish between worlds they've ruled out (in this case, w_2, w_3). Thus,

$$Pr(\mathtt{F} \mathrel{\Diamond\!\!\rightarrow} \bar{\mathtt{H}}) = 1$$

$$Pr(\mathtt{F} \Rightarrow \mathtt{H}) = 1/2$$

But now notice that if $Pr(\cdot|\mathtt{F} \Rightarrow \mathtt{H})$ encodes the probabilistic inferential dispositions of some agent, that agent must inferentially distinguish between worlds they've ruled out by assigning probability 0 to the sequence $\langle w_1, w_2, w_3\rangle$ and probability 1 to the sequence $\langle w_1, w_3, w_2\rangle$ even though they've ruled out both w_2, w_3. Since this is impossible, $Pr(\cdot|\mathtt{F} \Rightarrow \mathtt{H})$ does not encode the probabilistic inferential dispositions of any agent.

There is another interesting property that $Pr(\cdot|\mathtt{F} \Rightarrow \mathtt{H})$ lacks, and that is being normalized on a stable refined belief state:

A refined state $\mathtt{A}$ is stable iff $\Uparrow\Downarrow\mathtt{A} = \mathtt{A}$.

The refined state that $Pr(\cdot|\mathtt{F} \Rightarrow \mathtt{H})$ is normalized on is $\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow \mathtt{H} = \{\langle w_1, w_3, w_2\rangle\}$. But this state is not stable:

$$\Downarrow(\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow \mathtt{H}) = \{w_1\}$$

$$\Uparrow\Downarrow(\Uparrow\mathbf{BEL} \cap \mathtt{F} \Rightarrow \mathtt{H}) = \Uparrow\mathbf{BEL} = \left\{ \begin{array}{c} \langle w_1, w_2, w_3\rangle \\ \langle w_1, w_3, w_2\rangle \end{array} \right\}$$

Thus, we have seen exactly what feature of our theory motivates failures of Closure under Conditionalization—it is the thesis that an agent's inferential dispositions are determined by her factual beliefs in such a way so as not to distinguish between worlds the agent has ruled out. This allows us to avoid Santorio Triviality.

Having avoided two recent triviality results, we might feel content that our theory can predict plausible non-trivial probabilities for subjunctive conditionals. However, we have not proved this result. I turn to establishing this tenability result in the final section.

10.3 Subjunctive Tenability

As we have seen, subjunctive conditionals very often have modal bases that are not bounded by the relevant agent's belief state. Thus, our tenability result from Chapter 4, which assumed boundedness, will not immediately carry over to subjunctives. We need to extend our construction of Q, the measure over subsets of Ω, to all sequence sets, not just those giving permutations of **BEL**. Recall our definition of Q thus far:

(Q1) For any $w \in \mathbf{BEL}$:
$$Q(\Sigma[w^1]) = p(w^1)$$

(Q2) For any $w^1, w^2 \in \mathbf{BEL}$:
$$Q(\Sigma[w^1, w^2]) = \frac{p(w^2)}{P(\mathbf{BEL} \setminus \{w^1\})} \cdot p(w^1)$$

(Q3) For all $w^1, \ldots, w^n \in \mathbf{BEL}$:
$$Q(\Sigma[w^1, \ldots, w^n]) = \frac{p(w^n)}{P(\mathbf{BEL} \setminus \{w^1, \ldots, w^{n-1}\})} \cdot Q(\Sigma[w^1, \ldots, w^{n-1}])$$

To extend our construction outside of **BEL**, let $\Sigma[w^1, \ldots, w^n \parallel w^{n+1}, \ldots, w^{n+i}]$ be a sequence whose first n worlds are **BEL**-worlds (where $\{w^1, \ldots, w^n\} = \mathbf{BEL}$) and whose other i worlds are $\overline{\mathbf{BEL}}$-worlds. And let $\mathring{P}$ be an ur-probability function that assigns equal weight to every world.

(Q4) For all $w^1, \ldots, w^n \in \mathbf{BEL}$ and any $w^{n+1} \in \overline{\mathbf{BEL}}$:
$$Q(\Sigma[w^1, \ldots, w^n \parallel w^{n+1}]) = Q(\Sigma[w^1, \ldots, w^n]) \cdot \mathring{P}(\{w^{n+1}\}|\overline{\mathbf{BEL}})$$

This says that the probability of a sequence starting with $w^1, \ldots, w^n$, which exhausts the **BEL**-worlds, and continuing with w^{n+1} is the probability of the initial sequence times the ur-probability of selecting w^{n+1} from $\overline{\mathbf{BEL}}$.

(Q5) For all $w^1, \ldots, w^n \in \mathbf{BEL}$ and any $w^{n+1}, \ldots w^{n+i} \in \overline{\mathbf{BEL}}$:
$Q(\Sigma[w^1, \ldots, w^n \parallel w^{n+1}, \ldots, w^{n+i}]) = Q(\Sigma[w^1, \ldots, w^n \parallel w^{n+1}, \ldots, w^{n+i-1}]) \cdot \mathring{P}(\{w^{n+i}\}|\overline{\mathbf{BEL}} \setminus \{w^{n+1}, \ldots w^{n+i-1})$

This is the fully general definition of the probability of a sequence starting with $w^1, \ldots, w^n$, which exhausts the **BEL**-worlds, and continuing with $w^{n+1}, \ldots, w^{n+i}$.

This final definition is quite complicated, so it may help to look at an example.

- $\mathbf{W} = \{w_1, w_2, w_3, w_4, w_5\}$
- $\mathbf{BEL} = \{w_1, w_2\}$
- $p(w_1) = p(w_2) = 1/2$
- $\mathbf{A} = \{w_3, w_4\}, \mathbf{B} = \{w_3\}$
- $D^{\mathbf{A}}_{w_1} = \{w_3\}, D^{\mathbf{A}}_{w_2} = \{w_3, w_4\}$
- $\Uparrow\mathbf{BEL} = \left\{\begin{matrix} \sigma_1 : \langle w_1, w_2, w_3, w_4, w_5 \rangle & \sigma_7 : \langle w_2, w_1, w_3, w_4, w_5 \rangle \\ \sigma_2 : \langle w_1, w_2, w_3, w_5, w_4 \rangle & \sigma_8 : \langle w_2, w_1, w_3, w_5, w_4 \rangle \\ \sigma_3 : \langle w_1, w_2, w_4, w_3, w_5 \rangle & \sigma_9 : \langle w_2, w_1, w_4, w_3, w_5 \rangle \\ \sigma_4 : \langle w_1, w_2, w_4, w_5, w_3 \rangle & \sigma_{10} : \langle w_2, w_1, w_4, w_5, w_3 \rangle \\ \sigma_5 : \langle w_1, w_2, w_5, w_3, w_4 \rangle & \sigma_{11} : \langle w_2, w_1, w_5, w_3, w_4 \rangle \\ \sigma_6 : \langle w_1, w_2, w_5, w_4, w_3 \rangle & \sigma_{12} : \langle w_2, w_1, w_5, w_4, w_3 \rangle \end{matrix}\right\}$

Here are some probability assignments. You can verify that these correspond to the relevant proportion of $\sigma \in \Uparrow\mathbf{BEL}$.

$$Q(\Sigma[w_1]) = p(w_1) = 1/2$$

$$Q(\Sigma[w_1, w_2]) = \frac{p(w_2)}{P(\mathbf{BEL} - \{w_1\})} \cdot p(w) = {}^{1/2}/_{1/2} \cdot 1/2 = 1/2$$

$$Q(\Sigma[w_1, w_2 \parallel w_3]) = \underbrace{Q(\Sigma[w_1, w_2])}_{1/2} \cdot \underbrace{\mathring{P}(\{w_3\}|\overline{\mathbf{BEL}})}_{1/3} = 1/6$$

$$Q(\Sigma[w_1, w_2 \parallel w_3, w_4]) = \underbrace{Q(\Sigma[w_1, w_2 \parallel w_3])}_{1/6} \cdot \underbrace{\mathring{P}(\{w_4\}|\overline{\mathbf{BEL}} \setminus \{w_3\})}_{1/2} = 1/12$$

$$Q(\Sigma[w_1, w_2 \parallel w_3, w_4, w_5]) = \underbrace{Q(\Sigma[w_1, w_2 \parallel w_3, w_4])}_{1/12} \cdot \underbrace{\mathring{P}(\{w_5\}|\overline{\mathbf{BEL}} \setminus \{w_3, w_4\})}_{1} = 1/12$$

Notice that we predict $Pr(\mathtt{A} \Rightarrow \mathtt{B}) = \sum_x P(\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = x) \cdot x =$

- $P(\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = 1) = 1/2$
- $P(\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = 0.5) = 1/2$
- So, $\sum_x P(\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = x) \cdot x = 1 \cdot 1/2 + 1/2 \cdot 1/2 = 3/4$
  and
- $\mathtt{A} \Rightarrow \mathtt{B} = \{\sigma_1, \ldots, \sigma_6, \sigma_7, \sigma_8, \sigma_{11}\}$
- Thus, since each $\sigma \in \Uparrow\mathbf{BEL}$ has the same probability (= 1/12), we have $Pr(\mathtt{A} \Rightarrow \mathtt{B}) = 9 \cdot 1/12 = 9/12 = 3/4$.

We aim to prove that this holds generally (as discussed above in §9.1, I assume here that the subjunctive's domain is orthogonal to *Pr*):

> **Generalized Subjunctive Probability**
> For any *Pr* and non-conditional A, B, where $P([D^{\mathbf{A}} \neq \varnothing]) = 1$,
>
> $$Pr(\mathtt{A} \Rightarrow \mathtt{B}) = P(\mathbf{AB}) + \Big[\sum_x P([\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = x] \mid \bar{\mathbf{A}}) \cdot x\Big] \cdot P(\bar{\mathbf{A}})$$

We prove this by adapting the tenability result from Chapter 4. Given Strong Centering, Generalized Subjunctive Probability follows from:

> **Subjunctive Target**
>
> $$Pr(\mathtt{A} \Rightarrow \mathtt{B}|\bar{\mathtt{A}}) = \sum_x P([\mathring{P}(\mathbf{B}|D^{\mathbf{A}}_w) = x] \mid \bar{\mathbf{A}}) \cdot x$$

Again, I hold fixed t, m, g. Let $\Sigma[w]^{D^{\mathbf{A}}}_n$ be the set of w-led sequences whose first world in $D^{\mathbf{A}}_w$ is their nth, and let $\Sigma[w]^{D^{\mathbf{A}}_{\mathbf{B}}}_n$ be the set of

w-led sequences whose first world in $D_w^{\mathbf{A}}$ that is a **B**-world is their nth. Since $\mathtt{A} \Rightarrow \mathtt{B}$ is fully counterfactual for *Pr*, which is normalized on **BEL**, Subjunctive Target follows from:

The Key Subjunctive Lemma

$$\forall w \in \mathbf{BEL} \cap \bar{\mathbf{A}} : \forall n : |\mathbf{BEL}| < n \leq |\mathbf{W}| - |D_w^{\mathbf{A}}| + 1:$$
$$Q(\Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}} | \Sigma[w]_n^{D^{\mathbf{A}}}) = \mathring{P}(\mathbf{B}|D_w^{\mathbf{A}})$$

As before, Subjunctive Target follows from this because $\mathtt{A} \Rightarrow \mathtt{B}$ is the set of sequences whose first world in $D_w^{\mathbf{A}}$ is a **B**-world. The $\bar{\mathbf{A}}$-led sequences in $\Uparrow\mathbf{BEL}$ can be partitioned into n groups, according to whether their first world in $D_w^{\mathbf{A}}$ is their nth. Then, what the Key Subjunctive Lemma states is that, for each $\bar{\mathbf{A}}$-world w in **BEL** and each n, the probability of the set of sequences whose first $D_w^{\mathbf{A}}$ that is a **B**-world is their nth ($= \Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}}$) given the sequence set whose first world in $D_w^{\mathbf{A}}$ is their nth ($= \Sigma[w]_n^{D^{\mathbf{A}}}$) is equal to $\mathring{P}(\mathbf{B}|D_w^{\mathbf{A}})$. But if this holds for all possible places we might find the first $D_w^{\mathbf{A}}$-world in a sequence in $\Uparrow\mathbf{BEL}$ (that is, for all $n : |\mathbf{BEL}| < n \leq |\mathbf{W}| - |D_w^{\mathbf{A}}| + 1$), then it holds generally, and thus Target follows.

Proof. Let w be an arbitrary $\mathbf{BEL} \cap \bar{\mathbf{A}}$-world and n some number such that $|\mathbf{BEL}| < n \leq |\mathbf{W}| - |D_w^{\mathbf{A}}| + 1$. We assume $\mathtt{A} \Rightarrow \mathtt{B}$ is fully counterfactual for *Pr*. Start with the following facts:

(i) $$Q(\Sigma[w]_n^{D^{\mathbf{A}}}) = \sum_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum_{w_n \in D_w^{\mathbf{A}}} Q(\Sigma[w,\ldots \parallel \ldots w_{n-1}, w_n])$$
$$= \sum_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum_{w_n \in D_w^{\mathbf{A}}} Q(\Sigma[w,\ldots \parallel \ldots w_{n-1}]) \cdot \mathring{P}(\{w_n\}|\overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})$$

(ii) $$Q(\Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}}) = \sum_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum_{w_n \in D_w^{\mathbf{A}} \cap \mathbf{B}} Q(\Sigma[w,\ldots \parallel \ldots w_{n-1}, w_n])$$
$$= \sum_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum_{w_n \in D_w^{\mathbf{A}} \cap \mathbf{B}} Q(\Sigma[w,\ldots \parallel \ldots w_{n-1}]) \cdot \mathring{P}(\{w_n\}|\overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})$$

We have:

$$Q(\Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}} | \Sigma[w]_n^{D^{\mathbf{A}}}) = \frac{Q(\Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}} \cap \Sigma[w]_n^{D^{\mathbf{A}}})}{Q(\Sigma[w]_n^{D^{\mathbf{A}}})}$$

Thus,

$$\frac{Q(\Sigma[w]_n^{D_{\mathbf{B}}^{\mathbf{A}}} \cap \Sigma[w]_n^{D^{\mathbf{A}}})}{Q(\Sigma[w]_n^{D^{\mathbf{A}}})} =$$

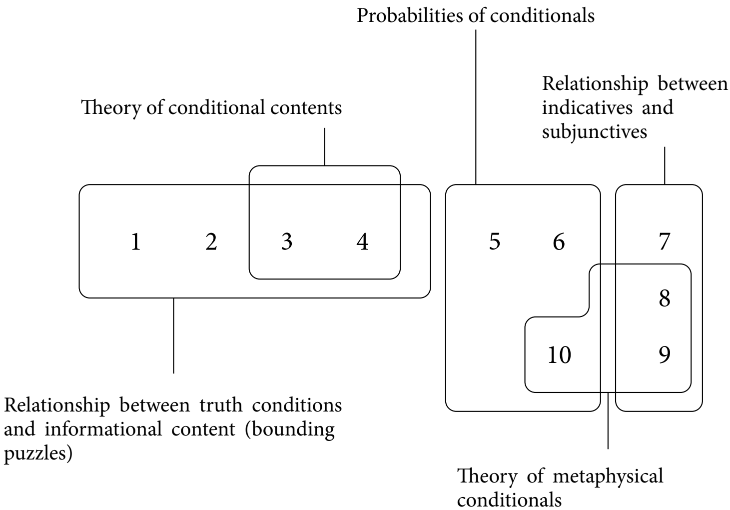
$$\frac{\sum\limits_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum\limits_{w_n \in D_w^{\mathbf{A}} \cap \mathbf{B}} Q(\Sigma[w, \ldots \parallel \ldots w_{n-1}]) \cdot \mathring{P}(\{w_n\} | \overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})}{\sum\limits_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum\limits_{w_n \in D_w^{\mathbf{A}}} Q(\Sigma[w, \ldots \parallel \ldots w_{n-1}]) \cdot \mathring{P}(\{w_n\} | \overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})} =$$

$$\frac{\sum\limits_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum\limits_{w_n \in D_w^{\mathbf{A}} \cap \mathbf{B}} \mathring{P}(\{w_n\} | \overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})}{\sum\limits_{w,\ldots,w_{n-1} \notin D_w^{\mathbf{A}}} \sum\limits_{w_n \in D_w^{\mathbf{A}}} \mathring{P}(\{w_n\} | \overline{\mathbf{BEL}} \setminus \{\ldots, w_{n-1}\})} =$$

$$\frac{\sum\limits_{w_n \in D_w^{\mathbf{A}} \cap \mathbf{B}} \mathring{P}(\{w_n\})}{\sum\limits_{w_n \in D_w^{\mathbf{A}}} \mathring{P}(\{w_n\})} = \mathring{P}(\mathbf{B} | D_w^{\mathbf{A}})$$

Conclusion

Here is a brief recap of the major points defended in the book. Recall the visual layout of the topics:



In Part I, I articulated, and attempted to solve, a puzzle about the relationship between the truth conditions of conditionals and their informational contents (what they communicate). My solution involved taking conditional contents to encode inferential dispositions and, hence, to be non-factual. In Part II, I extended this discussion to show how a factual probability measure determines a probability measure over non-factual conditional contents, encoding an agent's probabilistic inferential dispositions; the resulting view predicted the intuitively correct probabilities for conditionals while avoiding various triviality results. Finally, in Part III, I discussed the semantic and pragmatic relationship between indicative and subjunctive conditionals, articulating a uniform theory on which metaphysical subjunctive conditionals encode properties of sufficiency networks.

Having now come to the end of this journey, what issues remain on the horizon? Numerous projects involving applications of the strategies developed here suggest themselves. Here are a few:

- *Indeterminacy.* If I am right, conditionals are non-factual, and a distinctive mark of non-factuality, as compared with indeterminacy, is that we can know that a conditional $A \rightarrow B$ is non-factual throughout some possibly actual region of logical space, where this non-factuality is incorporated into our estimation of the probability of the conditional, resulting in a determinate credence (or belief) in the conditional's content. This differs, I think, from indeterminacy, whose presence percolates up to our attitudes targeting such expressions—for instance, we might think that the proper attitude to take towards borderline instances of some vague predicate is not partial belief but rather indeterminate belief (cf. Dorr 2003). A full comparative study of the similarities and differences between vague predicates (and other instances of indeterminacy, such as the indeterminacy exhibited by the open future) and the non-factuality of conditionals warrants further work.
- *Probability and non-factuality.* Similarly, the view I defend is a novel kind of non-factualism that allows for non-trivial assignments of probability to non-factual contents. This is a prediction that previous versions of nonfactualism have grappled with to varying degrees of success (see Yalcin 2007, 2012c; Moss 2015, 2018). If my view is ultimately successful, we then have a point of comparison and a possible strategy to pursue for expressivist theories in other domains, for instance, moral and aesthetic language, where the issue of uncertainty regarding such claims has been raised as a challenge (see Smith 2002, Lenman 2007, Sepielli 2012, Eriksson & Olinder 2016, Ridge 2018, Staffel 2019, Beddor 2020).
- *Middle knowledge.* A classic controversial issue in philosophy of religion is the question of whether God has "middle knowledge"

(*scientia media*)—knowledge of what people would freely do in every possible circumstance (the doctrine of Molinism, so called because of its early defender Luis de Molina—see Adams 1977, Flint 1988 and references therein). A standard argument against the possibility of middle knowledge is that such counterfactuals of freedom cannot be determinately true, and thus cannot be known in advance of the free choice of the creatures that make them true (see, for instance, Adams 1977, Hasker 1986, Van Inwagen 1997). One problem with this argument, as noted by Adams 1977: 114–115 (which he attributes to Suarez), is that such counterfactuals cannot thereby be possibly or probably true, contravening ordinary intuitions that Curley might freely accept the bribe if offered. The version of non-factualism motivated in this book provides the anti-Molinist with an answer to this challenge: these counterfactuals of freedom are unknowable (since only factual truths can be knowable) but can still be more or less likely true.

- *Compositional details.* Most of this project has been a discussion about issues connected to the relationship between the semantics of conditionals and the contents they express. However, in focusing on these topics, I have had to set aside a number of important questions about the compositional semantics of conditionals. To list a few outstanding issues worth exploring in future work:
  - Import export seems valid for indicative conditionals but not subjunctive conditionals. Here is a counterexample involving the latter from Steve Yablo (personal communication):

    (9) a. If Smith had been six feet tall, then if he had been a little taller, he would have been more than six feet tall.
    b. #If Smith had been six feet tall and a little taller, he would have been more than six feet tall.

  - Although I gave hints at some strategies to pursue, a plausible compositional semantics for a temporal past theory of subjunctive conditionals, including one that extends to counterfactual wishes, remains an open issue.

    - I have also completely set aside a wide range of conditionals—biscuit conditionals, conditional imperatives, conditional questions, etc. It remains to be seen whether the theory defended here can be integrated with plausible theories of these latter conditionals.
- *Backtracking and deliberation*. It matters to decision making how various outcomes objectively depend on our possible courses of action—this is why our probabilities in subjunctive conditionals are standardly thought to provide the weights on the values we assign to outcomes when rationally deciding what to do. At the same time, it's well known that some objective connections between our choices and certain outcomes do not matter to rational decision making—those that involve backtracking reasoning. My theory helps to illuminate why: backtracking interpretations of subjunctive conditionals involve intervening in a sufficiency network at a time that forces us to consider both upstream and downstream consequences of supposing its antecedent into that network, and it is the combination of generating changes throughout the network in both directions simultaneously that leads to the resultant outcomes of our choices being irrelevant to the rationality of that choice.

Bibliography

Abusch, Dorit. 1998. Generalizing Tense Semantics for Future Contexts. *Pages 13–35 of*: Rothstein, Susan (ed.), *Events and Grammar*. Dordrecht: Kluwer Academic.

Adams, Ernest. 1965. The Logic of Conditionals. *Inquiry*, **8**(1), 166–197.

Adams, Ernest. 1975. *The Logic of Conditionals*. Synthese Library, vol. 86. Boston: D. Reidel.

Adams, Robert Merrihew. 1977. Middle Knowledge and the Problem of Evil. *American Philosophical Quarterly*, **14**(2), 109–117.

Anderson, Alan Ross. 1951. A Note on Subjunctive and Counterfactual Conditionals. *Analysis*, **11**, 35–38.

Arregui, Ana. 2005. *On the Accessibility of Possible Worlds: The Role of Tense and Aspect*. Ph.D. thesis, University of Massachusetts at Amherst.

Arregui, Ana. 2007. When Aspect Matters: The Case of *Would* Conditionals. *Natural Language Semantics*, **15**, 221–264.

Arregui, Ana. 2009. On Similarity in Counterfactuals. *Linguistics and Philosophy*, **32**, 245–278.

Austin, J. L. 1970. Ifs and Cans. *Pages 153–180 of*: *Philosophical Papers*, 2nd ed. London: Oxford University Press.

Bacon, Andrew. 2015. Stalnaker's Thesis in Context. *The Review of Symbolic Logic*, **8**(1), 131–163.

Bar-Hillel, Maya. 1980. The Base-Rate Fallacy in Probability Judgments. *Acta Psychologica*, **44**, 211–233.

Barnes, Elizabeth, & Cameron, Ross. 2009. The Open Future: Bivalence, Determinism, and Ontology. *Philosophical Studies*, **146**, 291–309.

Barnes, Elizabeth, & Cameron, Ross P. 2011. Back to the Open Future. *Philosophical Perspectives*, **25**, 1–26.

Barnes, Elizabeth, & Williams, J. Robert G. 2011. A Theory of Metaphysical Indeterminacy. *Pages 103–148 of*: Bennett, Karen, and Zimmerman, Dean (eds.), *Oxford Studies in Metaphysics*, vol. 6. Oxford: Oxford University Press.

Baron, Sam, Colyvan, Mark, & Ripley, David. 2017. How mathematics can make a difference. *Philosophers' Imprint*, **17**(3), 1–19.

Beddor, Bob. 2020. Fallibility for expressivists. *Australasian Journal of Philosophy*, **98**(4), 763–777.

Bennett, Jonathan. 2003. *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Bjerring, Jens Christian. 2013. On Counterpossibles. *Philosophical Studies*, **168**(2), 327–353.

Blumberg, Kyle, & Lederman, Harvey. 2020. Revisionist reporting. *Philosophical Studies*, **178**(3), 755–783.

Bradley, Richard. 2000. A preservation condition for conditionals. *Analysis*, **60**(267), 219–222.

Bradley, Richard. 2002. Indicative conditionals. *Erkenntnis*, **56**(3), 345–378.

Bradley, Richard. 2007. A defence of the Ramsey test. *Mind*, **116**(461), 1–21.

Bradley, Richard. 2012. Multidimensional Possible-World Semantics for Conditionals. *Philosophical Review*, **121**(4), 539–571.

Briggs, Ray. 2012. Interventionist Counterfactuals. *Philosophical Studies*, **160**, 139–166.

Brogaard, Berit, & Salerno, Joe. 2013. Remarks on Counterpossibles. *Synthese*, **190**, 639–660.

Cariani, Fabrizio, & Santorio, Paolo. 2018. 'Will' Done Better: Selection Semantics, Future Credence, and Indeterminacy. *Mind*, **127**(505), 129–165.

Chakrabarti, Kisor Kumar. 2010. *Classical indian philosophy of induction: The Nyaya viewpoint*. Lexington Books.

Charlow, Nate. 2016. Triviality for Restrictor Conditionals. *Nous*, **50**(3), 553–564.

Charlow, Nate. 2019. The spectre of triviality. *Analysis*, **79**(4), 595–605.

Charlow, Nate. 2020. Grading modal judgment. *Mind*, **129**(515): 769–807.

Ciardelli, Ivano. 2021. The restrictor view, without covert modals. *Linguistics and Philosophy*.

Cohen, Daniel. 1988. The Problem of Counterpossibles. *Notre Dame Journal of Formal Logic*, **29**(1), 91–101.

Comrie, Bernard. 1985. *Tense*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Condoravdi, Cleo. 2002. Temporal Interpretation of Modals: Modals for the Present and for the Past. *Pages 59–87 of*: Beaver, David, Kaufmann, Stefan, Clark, Brady, & Casillas, Luis (eds.), *The Construction of Meaning*. Stanford, CA: CSLI.

Copley, Bridget. 2002. *The Semantics of the Future*. Ph.D. thesis, Massachussetts Institute of Technology, Cambridge.

Cross, Charles. 2009. Conditional Excluded Middle. *Erkenntnis*, **70**, 173–188.

DeRose, Keith. 1999. Can It Be That It Would Have Been Even Though It Might Not Have Been? *Philosophical Perspectives*, **13**, 385–413.

DeRose, Keith. 2010. The Conditionals of Deliberation. *Mind*, **119**(473), 1–42.

DeRose, Keith, & Grandy, Richard. 1999. Conditional Assertions and "Biscuit" Conditionals. *Nous*, **33**(3), 405–420.

Dorr, Cian. 2003. Vagueness without ignorance. *Philosophical Perspectives*, **17**(1), 83–113.

Dorr, Cian. 2016. Against Counterfactual Miracles. *Philosophical Review*, **125**(2), 241–286.

Douven, Igor. 2008. Kaufmann on the Probabilities of Conditionals. *Journal of Philosophical Logic*, **37**(3), 259–266.

Downing, P. B. 1959. Subjunctive Conditionals, Time Order, and Causation. *Meeting of the Aristotelian Society*, **59**, 125–140.

Dowty, David R. 1979. *Word Meaning and Montague Grammar*. London: Springer.

Dudman, V. H. 1983. Tense and Time in English Verb Clusters of the Primary Pattern. *Australasian Journal of Linguistics*, **3**, 25–44.

Dudman, V. H. 1984. Parsing 'If'-Sentences. *Analysis*, **44**(4), 145–153.

Dudman, V. H. 1988. Indicative and Subjunctive. *Analysis*, **48**(3), 113–122.

Ebert, Christian, Endriss, Cornelia, & Hinterwimmer, Stefan. 2008. A Unified Analysis of Indicative and Biscuit Conditionals as Topics. *Pages 266–283 of*: Friedman, T., & Ito, S. (eds.), *SALT XVIII*, Ithaca, NY: Cornell University.

Edgington, Dorothy. 1991. The Mystery of the Missng Matter of Fact. *Proceedings of the Aristotelian Society, Supplementary Volumes*, **65**, 185–209.

Edgington, Dorothy. 1995. On Conditionals. *Mind*, **104**, 235–329.

Edgington, Dorothy. 2004. Counterfactuals and the Benefit of Hindsight. *Pages 12–28 of*: *Cause and Chance: Causation in an Indeterministic World*. New York: Routledge.

Edgington, Dorothy. 2007. On Conditionals. *Pages 127–221 of*: Gabbay, D. M., & Guenthner, F. (eds.), *Handbook of Philosopical Logic*, vol. 14. New York: Springer.

Edgington, Dorothy. 2008. I–Counterfactuals. *Pages 1–21 of*: *Proceedings of the Aristotelian Society (Hardback)*, vol. 108. Wiley Online Library.

Edgington, Dorothy. 2014. Estimating Conditional Chances and Evaluating Counterfactuals. *Studia Logica*, **102**, 691–707.

Egan, Andy, Hawthorne, John, & Weatherson, Brian. 2005. Epistemic Modals in Context. *Pages 131–170 of*: Preyer, George, & Peter, George (eds.), *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Oxford University Press.

Eriksson, John, & Olinder, Ragnar Francén. 2016. Non-cognitivism and the classification account of moral uncertainty. *Australasian Journal of Philosophy*, **94**(4), 719–735.

von Fintel, Kai. 1994. *Restrictions on Quantifier Domains*. Ph.D. thesis, University of Massachusetts at Amherst.

von Fintel, Kai. 1997. The Presupposition of Subjunctive Conditionals. *In:* Percus, Orin, & Sauerland, Uri (eds.), *MIT Working Papers in Linguistics*, vol. 25. Cambridge, MA: Massachusetts Institute of Technology.

von Fintel, Kai. 2004. Would You Believe It? The King of France is Back! (Presuppositions and Truth-Value Intuitions). *Pages 315–342 of*: Reimer, Marga, & Bezuidenhout, Anne (eds.), *Descriptions and Beyond*. Oxford: Oxford University Press.

von Fintel, Kai. 2012a. *The best we can (expect to) get? Challenges to the classic semantics for deontic modals*. Manuscript in preparation.

von Fintel, Kai. 2012b. Subjunctive Conditionals. *Pages 466–477 of*: Russell, Gillian, & Fara, Delia Graff (eds.), *The Routledge Companion to Philosophy of Language*. New York: Routledge.

von Fintel, Kai, & Gillies, Anthony. 2008. CIA Leaks. *Philosophical Review*, **117**(1), 77–98.

von Fintel, Kai, & Gillies, Anthony. 2010. Must ... Stay ... Strong! *Natural Language Semantics*, **18**(4), 351–383.

von Fintel, Kai, & Gillies, Anthony. 2015. *Hedging Your 'Ifs' and Vice Versa*. Manuscript in preparation.

von Fintel, Kai, & Heim, Irene. 2012. *Intensional Semantics*. Lecture Notes.

von Fintel, Kai, & Iatridou, Sabine. 2002. *If and When If-Clauses can Restrict Quantifiers*. http://web.mit.edu/fintel/www/lpw.mich.pdf.

von Fintel, Kai, & Iatridou, Sabine. 2008. How to Say *Ought* in Foreign: The Composition of Weak Necessity Modals. *Pages 115–141 of*: Guéron, Jacqueline, & Lecarme, Jacqueline (eds.), *Time and Modality*. New York: Springer.

Fitelson, Branden. 2015. The Strongest Possible Lewisian Triviality Result. *Thought: A Journal of Philosophy*, **4**(2), 69–74.

Flint, Thomas. 1988. Two accounts of providence. *Pages 147–181 of*: Morris, Thomas V. (ed.), *Divine and Human Action*. Ithaca, NY: Cornell University Press.

Francez, Itamar. 2015. Chimerical conditionals. *Semantics and Pragmatics*, **8**, 1–35.

Franke, Michael. 2007. The Pragmatics of Biscuit Conditionals. *Pages 91–96 of*: Aloni, M., Dekker, P., and Roelofsen, F. (eds.), *Proceedings of the 16th Amsterdam*. ILLC/Department of Philosophy University of Amsterdam.

Gabaldon, Diana. 2005. *The Fiery Cross*. New York: Dell.

Gärdenfors, Peter. 1981. An epistemic approach to conditionals. *American philosophical quarterly*, **18**(3), 203–211.

Gärdenfors, Peter. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. Cambridge: Massachusetts Institute of Technology Press.

Gärdenfors, Peter. 2003. *Belief revision*. Vol. 29. New York: Cambridge University Press.

Gärdenfors, Peter, & Makinson, David. 1988. Revisions of knowledge systems using epistemic entrenchment. *Pages 83–95 of*: Vardi, Moshe (ed.), *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*. San Francisco, CA: Morgan Kaufmann.

Geurts, Bart. 2005. Entertaining Alternatives: Disjunctions as Modals. *Natural Language Semantics*, **13**, 383–410.

Gibbard, Allan. 1981. Two Recent Theories of Conditionals. *Pages 211–247 of*: Harper, William L., Stalnaker, Robert, & Pearce, Glenn (eds.), *Ifs*. Dordrecht: Reidel.

Gibbard, Allan, & Harper, William L. 1981. Counterfactuals and Two Kinds of Expected Utility. *Pages 153–190 of*: Harper, William L., Stalnaker, Robert, & Pearce, Glenn (eds.), *Ifs*. Dordrecht: Reidel.

Gigerenzer, Gerd, & Hoffrage, Ulrich. 1995. How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, **102**(4), 684–704.

Gillies, Anthony. 2004. Epistemic Conditionals and Conditional Epistemics. *Nous*, **38**, 585–616.

Gillies, Anthony. 2009. On Truth-Conditions for If (But not Quite Only If). *Philosophical Review*, **118**(3), 325–349.

Gillies, Anthony. 2010. Iffiness. *Semantics & Pragmatics*, **3**, 1–42.

Goldstein, Simon, & Santorio, Paolo. 2021. Probability for Epistemic Modalities. *Philosopher's Imprint*.

Goodman, Jeffrey. 2004. An Extended Lewis/Stalnaker Semantics and the New Problem of Counterpossibles. *Philosophical Papers*, **33**(1), 35–66.

Goodman, Nelson. 1947. The Problem of Counterfactual Conditionals. *Journal of Philosophy*, **44**(5), 113–128.

Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Groenendijk, Jeroen, & Stokhof, Martin. 1984. *Studies on the Semantics of Questions and Studies on the Semantics of Questions and Studies in the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.

Hájek, Alan. 1989. Probabilities of Conditionals—Revisited. *Journal of Philosophical Logic*, **18**, 423–428.

Hájek, Alan. 2011a. Conditional Probability. *Pages 99–135 of*: Bandyopadhyay, Prasanta, & Forster, Malcolm (eds.), *Handbook for Philosophy of Statistics*. New York: Elsevier.

Hájek, Alan. 2011b. *Most Counterfactuals Are False*. Manuscript in preparation.

Hájek, Alan. 2011c. Triviality Pursuit. *Topoi*, **30**, 3–15.

Hájek, Alan. 2012. The Fall of "Adams' Thesis"? *Journal of Logic, Language, and Information*, **21**, 145–161.

Hájek, Alan. 2014. Probabilities of counterfactuals and counterfactual probabilities. *Journal of Applied Logic*, **12**, 235–251.

Hájek, Alan. 2021. Contra Counterfactism. *Synthese*, **199**, 181–210.

Hall, Ned. 1994. Correcting the guide to objective chance. *Mind*, **103**(412), 505–517.

Hansson, Sven Ove. 2006. Logic of belief revision. *In: Stanford Encyclopedia of Philosophy*. Stanford, CA: Stanford University Press.

Hansson, Sven Ove. 2012. *A textbook of belief dynamics: Theory change and database updating*. New York: Springer.

Hasker, William. 1986. A refutation of middle knowledge. *Nous*, **20**(4), 545–557.

Hawthorne, John. 2005. Chance and counterfactuals. *Philosophy and Phenomenological Research*, **70**(2), 396–405.

Hawthorne, John, Rothschild, Daniel, & Spectre, Levi. 2016. Belief is weak. *Philosophical Studies*, **173**(5), 1393–1404.

Heim, Irene. 1990. E-Type Pronouns and Donkey Anaphora. *Linguistics and Philosophy*, **13**, 137–177.

Heim, Irene. 1991. Articles and definiteness. *Pages 487–535 of*: von Stechow A., & Wunderlich, D. (eds.), *Semantics: An international handbook of contemporary research*, Berlin: de Gruyter.

Hiddleston, Eric. 2005. A Causal Theory of Counterfactuals. *Nous*, **39**(4), 632–657.

Higginbotham, James. 1996. The Semantics of Questions. *Pages 361–383 of*: Lappin, Shalom (ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell.

Higginbotham, James. 2003. Conditionals and Compositionality. *Philosophical Perspectives: Language and Philosophical Linguistics*, **17**, 181–194.

Hintikka, Jaakko. 1961. Modality and Quantification. *Theoria*, **27**, 110–128.

Hoffrage, Ulrich, Lindsey, Samuel, Hertwig, Ralph, & Gigerenzer, Gerd. 2000. Communicating Statistical Information. *Science*, **290**(5500), 2261–2262.

Horn, Laurence. 1989. *A Natural History of Negation*. Chicago: IL: CSLI.

Horn, Laurence. 1993. Economy and Redundancy in a Dualistic Model of Natural Language. *Pages 33–72 of*: Shore, Susanna, & Vilkuna, Maria (eds.), *SKY 1993: Yearbook of the Linguistic Association of Finland*. Helsinki: Linguistic Association of Finland.

Hurford, James. 1974. Exclusive or Inclusive Disjunction. *Foundations of Language*, **11**(3), 409–411.

Iatridou, Sabine. 2000. The Grammatical Features of Counterfactuality. *Linguistic Inquiry*, **31**, 231–270.

Iatridou, Sabine. 2014. Grammar Matters. *In:* Walters, Lee, & Hawthorne, John (eds.), *Conditionals, Probability, and Paradox: Themes from the Philosophy of Dorothy Edgington*. Oxford: Oxford University Press.

Ippolito, Michela. 2003. Presuppositions and Implicatures in Counterfactuals. *Natural Language Semantics*, **11**, 145–186.

Ippolito, Michela. 2006. Semantic Composition and Presupposition Projection in Subjunctive Conditionals. *Linguistics and Philosophy*, **29**, 631–672.

Ippolito, Michela. 2013. *Subjunctive Conditionals: A Linguistic Analysis*. Linguistic Inquiry Monograph Series. Cambridge: Massachusetts Institute of Technology Press.

Isard, Steven. 1974. What Would You Have Done If... *Theoretical Linguistics*, **1**, 233–255.

Jackson, Frank. 1977. A Causal Theory of Counterfactuals. *Australasian Journal of Philosophy*, **55**, 3–21.

Jackson, Frank. 1979. On Assertion and Indicative Conditionals. *Philosophical Review*, **88**, 565–589.

Jackson, Frank. 1987. *Conditionals*. Oxford: Blackwell.

James, Deborah. 1982. Past Tense and the Hypothetical: A Cross-Linguistic Study. *Studies in Language*, **6**, 375–403.

Jeffrey, Richard. 1991. Matter-of-fact Conditionals. *Aristotelian Society: Supplemental Volume*, **65**, 161–183.

Jenny, Matthias. 2018. Counterpossibles in Science: The Case of Relative Computability. *Noûs*, **52**(3), 530–560.

Kadmon, Nirit. 1987. *On Unique and Non-Unique Reference and Asymmetric Quantification*. Ph.D. thesis, University of Massachusetts at Amherst.

Kahneman, Daniel, & Tversky, Amos. 1973. On the Psychology of Prediction. *Psychological Review*, **80**, 237–251.

Kahneman, Daniel, & Tversky, Amos. 1982. Evidence Impact of Base Rates. *Pages 153–160 of*: Kahneman, Daniel, Slovic, Paul, & Tversky, Amos (eds.), *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Katzir, Roni, & Singh, Raj. 2013. Hurford disjunctions: Embedded exhaustification and structural economy. *Pages 201–216 of*: Etxeberria, U., Fălăuş, A., Irurtzun, A., & Leferman, B. (eds.), *Proceedings of Sinn und Bedeutung*, vol. 18. Bayonne and Vitoria-Gasteiz.

Kaufmann, Stefan. 2004. Conditioning Against the Grain: Abduction and Indicative Conditionals. *Journal of Philosophical Logic*, **33**, 583–606.

Kaufmann, Stefan. 2005a. Conditional Predictions. *Linguistics and Philosophy*, **28**, 181–231.

Kaufmann, Stefan. 2005b. Conditional Truth and Future Reference. *Journal of Semantics*, **22**, 231–280.

Kaufmann, Stefan. 2009. Conditionals Right and Left: Probabilities for the Whole Family. *Journal of Philosophical Logic*, **38**, 1–53.

Kaufmann, Stefan. 2013. Causal Premise Semantics. *Cognitive Science*, **37**(6), 1136–1170.

Kaufmann, Stefan. 2015. Conditionals, conditional probabilities, and conditionalization. *Pages 71–94 of*: *Bayesian natural language semantics and pragmatics*. New York: Springer.

Kaufmann, Stefan. 2017. The limit assumption. *Semantics and Pragmatics*, **10**, 1–29.

Khoo, Justin. 2011. Operators or Restrictors? A Reply to Gillies. *Semantics & Pragmatics*, **4**(4), 1–25.

Khoo, Justin. 2013. A Note on Gibbard's Proof. *Philosophical Studies*, **166**(1), 153–164.

Khoo, Justin. 2015. On Indicative and Subjunctive Conditionals. *Philosophers' Imprint*, **15**(32), 1–40.

Khoo, Justin. 2016. Probability of Conditionals in Context. *Linguistics and Philosophy*, **39**(1), 1–43.

Khoo, Justin. 2017. Backtracking Counterfactuals Revisited. *Mind*, **126**(503), 841–910.

Khoo, Justin. 2020. Review of Schulz's *Counterfactuals and Probability*. *The Philosophical Review*, **2020**, 481–495.

Khoo, Justin. 2021. Coordinating *if*s. *Journal of Semantics*, **38**(2), 341–361.

Khoo, Justin, & Mandelkern, Matthew. 2019. Triviality results and the relationship between logical and natural languages. *Mind*, **128**(510), 485–526.

Khoo, Justin, & Santorio, Paolo. 2018. Probabilities of conditionals in modal semantics. *In: NASSLLI*. Pittsburgh, PA: Carnegie Mellon University Press.

Klinedinst, Nathan. 2010. Quantified Conditionals and Conditional Excluded Middle. *Journal of Semantics*, **27**(3), 1–22.

Kneale, William, & Kneale, Martha. 1962. *The Development of Logic*. Oxford: Clarendon Press.

Krakauer, Barak. 2012. *Counterpossibles*. Ph.D. thesis, University of Massachusetts at Amherst.

Kratzer, Angelika. 1977. What 'Must' and 'Can' Must and Can Mean. *Linguistics and Philosophy*, **1**, 337–355.

Kratzer, Angelika. 1981. The Notional Category of Modality. *Pages 38–74 of*: Eikmeyer, H. J., & Rieser, H. (eds.), *Words, Worlds, and Contexts. New Approaches in Words Semantics*. Berlin: de Gruyter.

Kratzer, Angelika. 1986. Conditionals. *Chicago Linguistics Society*, **22**(2), 1–15.

Kratzer, Angelika. 1991. Modality. *Chap. 23, pages 639–650 of*: von Stechow, Arnim, & Wunderlich, Dieter (eds.), *Handbuch Semantik*. Berlin and New York: de Gruyter.

Kratzer, Angelika. 2012. *Collected Papers on Modals and Conditionals*. Oxford: Oxford University Press.

Krifka, Manfred. 2014. Embedding illocutionary acts. *Pages 59–87 of*: *Recursion: Complexity in cognition*. New York: Springer.

Lange, Marc. 2000. *Natural Laws in Scientific Practice*. Oxford: Oxford University Press.

Lassiter, Daniel. 2020. What we can learn from how trivalent conditionals avoid triviality. *Inquiry*.vol. 63(9–10): 1087–1114.

Leahy, Brian. 2011. Presuppositions and Antipresuppositions in Conditionals. *Pages 257–274 of*: Ashton, Neil, Chereches, Anca, & Lutz, David (eds.), *Proceedings of SALT 21*. CLC.

Leahy, Brian. 2016. On presuppositional implicatures. *Topoi*, **35**(1), 83–91.

Leahy, Brian. 2018. Counterfactual antecedent falsity and the epistemic sensitivity of counterfactuals. *Philosophical Studies*, **175**(1), 45–69.

Lenman, James. 2007. Expressivism and epistemology: What is moral inquiry? *Pages 63–81 of*: *Aristotelian Society Supplementary Volume*, vol. 81. Wiley Online Library.

Levi, Isaac. 1978. Subjunctives, dispositions and chances. *Pages 303–335 of*: *Dispositions*. New York: Springer.

Lewis, David. 1973a. *Counterfactuals*. Oxford: Blackwell.

Lewis, David. 1973b. Counterfactuals and Comparative Possibility. *Journal of Philosophical Logic*, **2**, 418–446.

Lewis, David. 1975. Adverbs of Quantification. *Pages 178–188 of*: Keenan, Edward L. (ed.), *Formal Semantics of Natural Language*. Cambridge, UK: Cambridge University Press.

Lewis, David. 1976. Probabilities of Conditionals and Conditional Probabilities. *The Philosophical Review*, **85**, 297–315.

Lewis, David. 1979a. Counterfactual Dependence and Time's Arrow. *Nous*, **13**, 455–476.

Lewis, David. 1979b. A Problem About Permission. *Pages 20–33 of*: *Papers in Ethics and Social Philosophy*. Cambridge, UK: Cambridge University Press.

Lewis, David. 1979c. Scorekeeping in a Language Game. *Journal of Philosophical Logic*, **8**, 339–359.

Lewis, David. 1980a. Index, Context, and Content. *Pages 79–100 of*: Kanger, Stig, & Ohman, Sven (eds.), *Philosophy and Grammar*. Dordrecht: Reidel.

Lewis, David. 1980b. A Subjectivist's Guide to Objective Chance. *Pages 263–293 of*: Jeffrey, Richard (ed.), *Studies in Inductive Logic and Probability*. University of California Press.

Lewis, David. 1981. Causal Decision Theory. *Australasian Journal of Philosophy*, **59**, 5–30.

Lewis, David. 1986a. Causation. *Pages 159–172 of*: *Philosophical Papers, Vol. 1*. New York: Oxford University Press.

Lewis, David. 1986b. Postscripts to "Counterfactual Dependence and Time's Arrow." *Pages 52–66 of*: *Philosophical Papers, Vol. 2*. Oxford: Oxford University Press.

Lewis, David. 1997. Finkish Dispositions. *The Philosophical Quarterly*, **47**(187), 143–158.

Lewis, David. 1999. Why conditionalize? *In: Papers in Metaphysics and Epistemology*. Cambridge, UK: Cambridge University Press.

Lewis, Karen S. 2016. Elusive counterfactuals. *Nous*, **50**(2), 286–313.

Liu, Fenrong, & Zhang, Jialong. 2010. New perspectives on moist logic. *Journal of Chinese Philosophy*, **37**(4), 605–621.

Lyons, John. 1977. *Semantics*. Cambridge, UK: Cambridge University Press.

Mackay, John. 2013. Quantifying over Possibilities. *Philosophical Review*, **122**(4), 577–617.

Mackay, John. 2015. Actuality and Fake Tense in Conditionals. *Semantics & Pragmatics*, **8**(12), 1–12.

Mackay, John. 2017. Past Tense and Past Times in Subjunctive Conditionals. *Pacific Philosophical Quarterly*, **98**, 520–535.

Mandelkern, Matthew. 2018. Talking about worlds. *Philosophical Perspectives*, **32**(1), 298–325.

Mandelkern, Matthew. 2020. How to do things with modals. *Mind & Language*, **35**(1), 115–138.

Mandelkern, Matthew, & Khoo, Justin. 2019. Against Preservation. *Analysis*, **79**(3), 424–436.

Martin, Charles B. 1994. Dispositions and conditionals. *The Philosophical Quarterly (1950–)*, **44**(174), 1–8.

McDermott, Michael. 2007. True Antecedents. *Acta Analytica*, **22**, 333–335.

McGee, Vann. 1985. A Counterexample to Modus Ponens. *The Journal of Philosophy*, **82**(9), 462–471.

McGee, Vann. 1989. Conditional Probabilities and Compounds of Conditionals. *Philosophical Review*, **98**, 485–542.

McGee, Vann. 2000. To Tell the Truth about Conditionals. *Analysis*, **60**(1), 107–111.

Mellor, D. H. 1993. How to believe a conditional. *The Journal of Philosophy*, **90**(5), 233–248.

Milne, Peter. 2003. The simplest Lewis-style triviality proof yet? *Analysis*, **63**(280), 300–303.

Moss, Sarah. 2013. Subjunctive Credences and Semantic Humility. *Philosophy and Phenomenological Research*, **87**(2), 251–278.

Moss, Sarah. 2015. On the semantics and pragmatics of epistemic vocabulary. *Semantics & Pragmatics*, **8**(5), 1–81.

Moss, Sarah. 2018. *Probabilistic Knowledge*. Oxford: Oxford University Press.

Nolan, Daniel. 1997. Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, **38**(4), 535–571.

Nozick, Robert. 1969. Newcomb's problem and two principles of choice. *Pages 114–146 of*: Rescher, Nicholas (ed.), *Essays in honor of Carl G. Hempel*. New York: Springer.

Ogihara, Toshiyuki. 2007. Tense and Aspect in Truth-Conditional Semantics. *Lingua*, **117**(2), 392–418.

Osborn, Jane M. 1965. Austin's Non-Conditional Ifs. *Journal of Philosophy*, **62**(23), 711–715.

Partee, Barbara. 1973. Some Structural Analogies between Tenses and Pronouns in English. *Journal of Philosophy*, **70**(18), 601–609.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Pollock, John L. 1976. *Subjunctive Reasoning*. New York: Springer.

Predelli, Stefano. 2009. Towards a Semantics for Biscuit Conditionals. *Philosophical Studies*, **142**(3), 293–305.

Pruss, Alexander. 2010. Probability and the Open Future View. *Faith and Philosophy*, **27**(2), 190–196.

Ramsey, Frank P. 1931. *The Foundations of Mathematics and other Logical Essays*. London: Kegan Paul, Trench, Trubner & Co.

Ridge, Michael. 2018. Normative certitude for expressivists. *Synthese*, **197**(8), 1–23.

Rieger, Adam. 2006. A Simple Theory of Conditionals. *Analysis*, **66**(3), 233–240.

Rips, Lance J. 2010. Two causal theories of counterfactual conditionals. *Cognitive science*, **34**(2), 175–221.

Rips, Lance J., & Edwards, Brian J. 2013. Inference and explanation in counterfactual reasoning. *Cognitive Science*, **37**(6), 1107–1135.

Roberts, Craige. 1996. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *In:* Yoon, Jae Hak, & Kathol, Andreas (eds.), *Ohio State University Working Papers in Linguistics*, vol. 49.

Roberts, Craige. 2012a. Information Structure: Afterword. *Semantics & Pragmatics*, **5**(7), 1–19.

Roberts, Craige. 2012b. Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics. *Semantics & Pragmatics*, **5**(6), 1–69.

Romero, Maribel. 2014. 'Fake Tense' in Counterfactuals: A Temporal Remoteness Approach. *Pages 47–63 of:* Crnič, Luka, & Sauerland, Uli (eds.), *The Art and Craft of Semantics: A Festschrift for Irene Heim*, vol. 2. Cambridge: MA: MITWPL.

Rothschild, Daniel. 2013. Do Indicative Conditionals Express Propositions? *Nous*, **47**(1), 49–68.

Rothschild, Daniel. 2014. Capturing the relationship between conditionals and conditional probability with a trivalent semantics. *Journal of Applied Non-Classical Logics*, **24**(1–2), 144–152.

Rothschild, Daniel. 2019. What it Takes to Believe. *Philosophical Studies*, 1345–1362.

Russell, Bertrand. 1905. On Denoting. *Mind*, **14**(56), 479–493.

Russell, Jeffrey Sanford, & Hawthorne, John. 2016. General Dynamic Triviality Theorems. *Philosophical Review*, **125**(3), 307–339.

Santorio, Paolo. 2012. Reference and Monstrosity. *The Philosophical Review*.

Santorio, Paolo. 2019. Interventions in premise semantics. *Philosophers' Imprint*, **19**(1), 1–27.

Santorio, Paolo. 2021. General Triviality for Counterfactuals. *Analysis*.

Santorio, Paolo. Forthcoming. Path Semantics for Indicative Conditionals. *Mind*.

Santorio, Paolo, & Goldstein, Simon. 2021. Probabilities for Epistemic Modalities. *Philosopher's Imprint*.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. *Linguistics and philosophy*, **27**(3), 367–391.

Sauerland, Uli. 2008. Implicated presuppositions. *The discourse potential of underspecified structures*, **8**, 581–600.

Schaffer, Jonathan. 2004. Counterfactuals, Causal Independence, and Conceptual Circularity. *Analysis*, **64**(4), 299–309.

Schaffer, Jonathan. 2009. On What Grounds What. *Chap. 12, pages 347–383 of:* Chalmers, David, & Manley, David (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Schaffer, Jonathan. 2016. Grounding in the image of causation. *Philosophical studies*, **173**(1), 49–100.

Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and philosophy*, **26**(1), 29–120.

Schlenker, Philippe. 2012. Maximize presupposition and Gricean reasoning. *Natural Language Semantics*, **20**(4), 391–429.

Schultheis, Ginger. 2021. *Counterfactual Probability*. Submitted for publication.

Schulz, Katrin. 2007. *Minimal Models in Semantics and Pragmatics: Free Choice, Exhaustivity, and Conditionals*. Ph.D. thesis, University of Amsterdam.

Schulz, Katrin. 2011. If you'd wiggled A, then B would've changed. *Synthese*, **179**(2), 239–251.

Schulz, Katrin. 2014. Fake Tense in Conditional Sentences: A Modal Approach. *Natural Language Semantics*, **22**(2), 117–144.

Schulz, Moritz. 2017. *Counterfactuals and probability*. Oxford: Oxford University Press.

Sellars, Wilfrid. 1962. Philosophy and the scientific image of man. *Pages 35–78 of:* Colodny, Robert (ed.), *Frontiers of Science and Philosophy*. University of Pittsburgh Press.

Sepielli, Andrew. 2012. Normative uncertainty for non-cognitivists. *Philosophical Studies*, **160**(2), 191–207.

Sider, Theodore. 2013. *Writing the Book of the World*. New York: Oxford University Press.

Siegel, Muffy E. 2006. Biscuit Conditionals: Quantification over Potential Literal Acts. *Linguistics and Philosophy*, **29**, 167–203.

Skyrms, Brian. 1981. The prior propensity account of subjunctive conditionals. *Pages 259–265 of*: Harper, William L., Stalnaker, Robert, & Pearce, Glenn (eds.), *Ifs*. New York: Springer.

Slote, Michael. 1978. Time in Counterfactuals. *The Philosophical Review*, **87**(1), 3–27.

Smith, Michael. 2002. Evaluation, uncertainty and motivation. *Ethical Theory and Moral Practice*, **5**(3), 305–320.

Snider, Todd, & Bjorndahl, Adam. 2015. Informative counterfactuals. *Pages 1–17 of*: D'Antonio, Sarah, Moroney, Mary, & Little, Carol Rose (eds.), *Semantics and Linguistic Theory*, vol. 25. Palo Alto, CA: Stanford.

Soames, Scott. 1982. How Presuppositions are Inherited: A Solution to the Projection Problem. *Linguistic Inquiry*, **13**(3), 483–545.

Staffel, Julia. 2019. Expressivism, Normative Uncertainty, and Arguments for Probabilism. *Oxford Studies in Epistemology*, **6**, 161–189.

Stalnaker, Robert. 1968. A Theory of Conditionals. *Pages 98–112 of*: Rescher, N. (ed.), *Studies in Logical Theory*. Oxford: Oxford University Press.

Stalnaker, Robert. 1970. Probability and Conditionals. *Philosophy of Science*, **37**(1), 64–80.

Stalnaker, Robert. 1972. Letter to David Lewis. *Pages 151–152 of*: Harper, William, Stalnaker, Robert, & Pearce, Glenn (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Dordrecht: Reidel.

Stalnaker, Robert. 1975. Indicative Conditionals. *Philosophia*, **5**, 269–286.

Stalnaker, Robert. 1976. Letter to van Fraassen. *Pages 302–306 of*: Harper, W., & Hooker, C. (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, vol. 1. Dordrecht: Reidel.

Stalnaker, Robert. 1978. Assertion. *Pages 315–332 of*: Cole, P. (ed.), *Syntax and Semantics 9: Pragmatics*. New York: Academic Press.

Stalnaker, Robert. 1980. A Defense of Conditional Excluded Middle. *Pages 87–104 of*: Harper, William L., Pearce, Glenn, & Stalnaker, Robert (eds.), *Ifs*. Dordrecht: Reidel.

Stalnaker, Robert. 1984. *Inquiry*. Cambridge: Massachusetts Institute of Technology Press.

Stalnaker, Robert. 1998. On the Representation of Context. *Journal of Logic, Language, and Information*, **7**, 3–19.

Stalnaker, Robert. 2011. Conditional Propositions and Conditional Assertions. *Pages 227–248 of*: Egan, Andy, & Weatherson, Brian (eds.), *Epistemic Modality*. Oxford: Oxford University Press.

Stalnaker, Robert. 2014. *Context*. Oxford: Oxford University Press.

Stalnaker, Robert. 2019. Counterfactuals and probability. *Chap. 11, pages 182–202 of*: *Knowledge and Conditionals: Essays on the Structure of Inquiry*. Oxford: Oxford University Press.

Stalnaker, Robert, & Jeffrey, Richard. 1994. Conditionals as Random Variables. *Pages 31–46 of*: Eells, Ellery, & Skyrms, Brian (eds.), *Probabilities* *and Conditionals: Belief Revision and Rational Decision*. Cambridge, UK: Cambridge University Press.

Starr, William B. 2013. A Uniform Theory of Conditionals. *Journal of Philosophical Logic*, **43**(6), 1019–1064.

Starr, William B. 2021. *Indicative Conditionals, Strictly*. Manuscript in preparation.

Steele, Susan. 1975. Past and Irrealis: Just what does it all mean? *International Journal of American Linguistics*, **41**(3), 200–217.

Stojnić, Una. 2018. Discourse, context and coherence: The grammar of prominence. *Pages 171–186 of*: Preyer, Gerhard (ed.), *Beyond Semantics and Pragmatics*. New York: Oxford University Press.

Stojnić, Una, Stone, Matthew, & Lepore, Ernest. 2018. Discourse and Logical Form: Pronouns, Attention and Coherence. *Lingusitics and Philosophy*, **40**(5), 519–547.

Strawson, Peter F. 1950. On referring. *Mind*, **59**(235), 320–344.

Street, Sharon. 2008. Constructivism about reasons. *Oxford studies in metaethics*, **3**(1), 207–245.

Sud, Rohan. 2018. Plurivaluationism, supersententialism and the problem of the many languages. *Synthese*, **197**(4), 697–1723.

Swanson, Eric. 2011a. Conditional Excluded Middle Without the Limit Assumption. *Philosophy and Phenomenological Research*, **85**(2), 301–321.

Swanson, Eric. 2011b. Subjunctive Biscuit and Stand-Off Conditionals. *Philosophical Studies*, **163**(3), 637–648.

Tedeschi, Philip. 1981. Some Evidence for a Branching-Futures Semantic Model. *Pages 239–269 of*: Tedeschi, P., & Zaenen, A. (eds.), *Syntax and Semantics: Tense and Aspect*, vol. 14. New York: Academic Press.

Teller, Paul. 1973. Conditionalization and Observation. *Synthese*, **26**, 218–258.

Thomason, Richmond, & Gupta, Anil. 1980. A Theory of Conditionals in the Context of Branching Time. *The Philosophical Review*, **89**(1), 65–90.

Tichý, Pavel. 1976. A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals. *Philosophical Studies*, **29**, 271–273.

Tooley, Michael. 2002. Backward Causation and the Stalnaker-Lewis Approach to Counterfactuals. *Analysis*, **62**(3), 191–197.

van Fraassen, Bas. 1976. Probabilities of Conditionals. *Pages 261–300 of*: Harper, William L., & Hooker, C. (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht: Reidel.

van Fraassen, Bas. 1980. Review of Brian Ellis, Rational Belief Systems. *Canadian Journal of Philosophy*, **10**, 497–511.

van Fraassen, Bas. 1981. Essences and laws of Nature. *Pages 189–200 of*: Healey, Richard (ed.), *Reduction, Time and Reality*. Cambridge, UK: Cambridge University Press.

Van Inwagen, Peter. 1997. Against middle knowledge. *Midwest Studies in Philosophy*, **21**(1), 225–236.

Van Rooij, Robert, & Schulz, Katrin. 2004. Exhaustive interpretation of complex sentences. *Journal of logic, language and information*, **13**(4), 491–519.

Vander Laan, David. 2004. Counterpossibles and similarity. *Pages 258–276 of*: Jackson, Frank, & Priest, Graham (eds.), *Lewisian Themes: The Philosophy of David K. Lewis*. Oxford: Clarendon.

Veltman, Frank. 2005. Making Counterfactual Assumptions. *Journal of Semantics*, **22**, 159–180.

Weatherson, Brian. 2001. Indicative and Subjunctive Conditionals. *The Philosophical Quarterly*, **51**(203), 200–216.

Williams, J. Robert G. 2008a. Chances, counterfactuals and similarity. *Philosophy and Phenomenological Research*, **77**(2), 385–420.

Williams, J. Robert G. 2008b. Conversation and Conditionals. *Philosophical Studies*, **138**, 211–223.

Williams, J. Robert G. 2008c. Multiple Actualities and Ontically Vague Identity. *Philosophical Quarterly*, **58**, 134–154.

Williams, J. Robert G. 2010. Defending Conditional Excluded Middle. *Nous*, **44**(4), 650–668.

Williams, J. Robert G. 2012. Counterfactual triviality: A Lewis-impossibility proof for counterfactuals. *Philosophy and Phenomenological Research*, **85**(3), 648–670.

Williams, J. Robert G. 2014. Decision making under indeterminacy. *Philosopher's Imprint*, **14**(4), 1–34.

Williamson, Timothy. 2015. Counterpossibles. *Topoi*, **37**(3), 357–368.

Yalcin, Seth. 2007. Epistemic Modals. *Mind*, **116**(464), 983–1026.

Yalcin, Seth. 2010. Probability Operators. *Philosophy Compass*, **5**(11), 916–937.

Yalcin, Seth. 2011. Nonfactualism About Epistemic Modality. *Pages 295–332 of*: Egan, Andy, & Weatherson, Brian (eds.), *Epistemic Modals*. Oxford: Oxford University Press.

Yalcin, Seth. 2012a. Bayesian Expressivism. *Pages 123–160 of*: Yalcin, Seth. (ed.), *Proceedings of the Aristotelian Society*. Vol. 112. No. 2_pt_2. Oxford, UK: Oxford University Press.

Yalcin, Seth. 2012b. Context Probabilism. *Pages 12–21 of*: *Proceedings of the 18th Amsterdam Colloquium*. New York: Springer.

Yalcin, Seth. 2012c. A counterexample to modus tollens. *Journal of Philosophical Logic*, **41**, 1001–1024.

Yalcin, Seth. 2018. *Conditional Belief and Conditional Assertion*. Manuscript in preparation.

Zimmerman, Thomas Ede. 2000. Free Choice Disjunction and Epistemic Possibility. *Natural Language Semantics*, **8**, 255–290.

Index of Names

Index of Terms