Coordination in Conversation
by
Matthew Mandelkern
B.A., University of Chicago (2011)
Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2017
© Massachusetts Institute of Technology 2017. All rights reserved.

Signature redacted

Author

Department of Linguistics and Philosophy
May 17, 2017

Signature redacted

Certified by

Robert Stalnaker
Laurance S. Rockefeller Professor of Philosophy
Thesis Supervisor

Certified by

Kai von Fintel
Andrew W. Mellon Professor of Linguistics
Thesis Supervisor

Signature redacted

Accepted by

Roger White
Professor of Philosophy
Chair of the Committee on Graduate Students
Coordination in Conversation
by
Matthew Mandelkern

Submitted to the Department of Linguistics and Philosophy
on May 17, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Philosophy

Abstract

I give an account of the meaning of epistemic modals—words like ‘might’ and ‘must’, on a broadly epistemic interpretation—and how speakers use them to coordinate on their information.

I begin by exploring what epistemic modals mean. Motivated by embedding data which are problematic for almost all existing accounts, I develop a new semantics for epistemic modals which I call the bounded theory. The bounded theory comprises a standard relational semantics together with a constraint which entails that local information is always taken into account in the evaluation of epistemic modals. I argue that the bounded theory makes sense of the subtle embedding behavior of epistemic modals—thus providing an adequate account of their meaning—and sheds new light on the way that local information is structured in natural language.

In Chapter 2, I turn to the question of how speakers use epistemic modals to coordinate on their common information. I argue that we can make sense of the characteristic dynamics of epistemic modality within a relational semantic framework (like the one given in Chapter 1), together with a contextualist approach to the dynamics of conversation. The key to doing so is to take the relation in question to be determined by the interlocutors’ prospective common attitudes.

Chapter 3 explores the contribution of ‘must’. Assertions of ‘Must p’ and assertions of p alone seem to have the same basic goal, and yet their felicity conditions subtly differ. I argue for a new characterization of those differences, and a broadly pragmatic explanation of those differences based on the approach to epistemic modals advocated in the preceding chapters.

In Chapter 4, I return to semantic questions. I develop formal tools to compare the expressive power of semantic theories with respect to the embedding data which they can account for. These tools show that the relational theory can account for all the embedding data which recent revisionary theories can account for, but not vice versa, a fact which necessitates a substantial shift in perspective on the debate regarding the semantics of epistemic modals.

Thesis Supervisor: Robert Stalnaker
Title: Laurance S. Rockefeller Professor of Philosophy

Thesis Supervisor: Kai von Fintel
Title: Andrew W. Mellon Professor of Linguistics
Acknowledgments

I am incredibly grateful to my committee for the countless meetings, informal discussions, seminars, and email exchanges which, more than anything else, have taught me how to work on philosophy of language, semantics, and pragmatics. I was drawn into philosophy of language by reading Context and Content, and it has been an extraordinary privilege to study with Robert Stalnaker in the years since; he has been an inspiration and model in every way, and will continue to be throughout my career. I have never felt confident in a paper until it could pass through the alembic of Kai von Fintel’s scrutiny; seminars and conversations with Kai have been a source of unbridled intellectual joy, and the starting point for much of the work in this dissertation. Meetings with Stephen Yablo have been a boundless source of freewheeling intellectual stimulation, teaching me the connection of philosophy of language to just about everything, and the essential role that humor plays in good philosophy. I have had few experiences as challenging and exciting as Irene Heim’s seminars; from them, and from conversation with Irene, I have learned to never fail to be puzzled by language, and to never be satisfied with the limitations of anyone’s theory of what is going on. Finally, it was a great stroke of luck for me that, in my second year, MIT hired Justin Khoo; he has been a wonderful mentor and guide through graduate school and philosophy of language, and I am proud that I will be able to say that I was his first PhD student—one of the rare accolades of which there is little risk of either disputation or deprivation.

In addition to my committee, I owe great debts to many informal advisors. Daniel Rothschild has been an incredible mentor and interlocutor, from whom I have learned an enormous amount about nearly everything that I have worked on in graduate school. Regular discussions with Dilip Ninan over coffee have been one of the most pleasant and stimulating fixtures of graduate school. Discussions and email exchanges with Fabrizio Cariani, Angelika Kratzer, and Paolo Santorio have played formative roles in my thinking about modality.

My peers and collaborators have played a formative role in my education. Vera Flocke has supported and enriched my intellectual and personal life since we met in Berlin. Rose Lenehan’s friendship and intellectual comradeship have been a joy at MIT. Sophie Horowitz and Arden Ali helped me navigate graduate school from my earliest days. I have learned enormously from talking with my officemates David Boylan, Cosmo Grant, Matthias Jenny, Milo Phillips-Brown, and Ginger Schultheis—who have made grad school much more fun than it should have been—as well as Nilanjan Das, Maša Močnik, Jacopo Romoli, and Bernhard Salow. I doubt I would be in philosophy today if not for my undergraduate thesis adviser, Agnes Callard; I have continued to learn from discussions with her in the years since. Joshua Knobe and Jonathan Phillips have expanded my intellectual horizons by introducing me to experimental methodologies and topics in moral psychology. MIT has been a wonderful place to work on topics at the intersection of linguistics and philosophy, and I am grateful for the many environments which have made interdisciplinary work
possible, in particular the Linguistics and Philosophy Reading Group. Before Bob, I had Richard Holton, and then Julia Markovits, as wonderful advisers (I am sorry to say that I apparently caused both sufficient distress that they left the Institute within a year of taking on that role; I wish I could report that the pattern was not repeated a third time, but Bob’s retirement in my fourth year makes that a partial truth at best). I have also learned a great deal from Alex Byrne, Andy Egan, Danny Fox, Sally Haslanger, Sabine Iatridou, Rae Langton, Vann McGee, Agustín Rayo, Kieran Setiya, Brad Skow, Jack Spencer, and Roger White. I have greatly benefited from discussions with Daniel Altshuler, Ashley Atkins, Dan Baras, Chris Barker, Justin Bledin, Jennifer Carr, Sam Carter, Simon Charlow, Cleo Condoravdi, Josh Dever, Ryan Doody, Kevin Dorst, Janice Dowell, Daniel Drucker, Peter van Elswyk, Branden Fitelson, Peter Fritz, Simon Goldstein, Valentine Hacquard, Jay Hodges, Hans Kamp, Brendan de Kennessey, Daniel Lassiter, Harvey Lederman, Pengbo Liu, Jack Marley-Payne, Sarah Moss, Sarah Murray, Kevin Richardson, Damien Rochford, Philippe Schlenker, Anders Schoubye, Brett Sherman, Florian Schwarz, Alex Silk, Daniel Skibra, Shane Steinert-Threlkeld, Eric Swanson, Zoltán Szabó, Wataru Uegaki, Brian Weatherson, Seth Yalcin, Martín Abreu Zavaleta, and Jérémie Zehr.

I thank, finally, my family and my friends outside of philosophy. I was fortunate that one of oldest friends, Tim, ended up in Boston; we have had many wonderful runs and evenings cooking together with Christi, and, more recently, with Danielle and Sam. And Paul, Michael, Sean, Ben and Anastasia, my parents, and Raphaël have been sources of great joy and support over the past five years.
# Contents

Introduction ............................................................................................................. 11

1 Bounded modality .................................................................................................. 15
  1.1 Introduction ........................................................................................................ 15
  1.2 The classical strategy ....................................................................................... 18
  1.3 Disjunctions ....................................................................................................... 20
  1.4 Bounded modality ............................................................................................. 29
    1.4.1 Local contexts: The basics ...................................................................... 30
    1.4.2 Local relational semantics ...................................................................... 31
    1.4.3 Local contexts, more precisely ............................................................. 33
  1.5 Explaining Wittgenstein disjunctions .............................................................. 35
  1.6 Order and the dynamics of information ......................................................... 38
  1.7 Empirical reach ................................................................................................ 44
    1.7.1 Attitudes and conditionals ...................................................................... 44
    1.7.2 Quantifiers ............................................................................................... 46
    1.7.3 Non-default readings .............................................................................. 52
  1.8 Alternative solutions ....................................................................................... 53
    1.8.1 Enriched update semantics (i): Yalcin 2015 ....................................... 53
    1.8.2 Enriched update semantics (ii): Rothschild and Klinedinst 2015 ....... 55
    1.8.3 Probabilistic contents: Moss 2015 ......................................................... 57
  1.9 Conclusion ......................................................................................................... 59

2 How to do things with modalities ............................................................................ 63
  2.1 Introduction ........................................................................................................ 63
  2.2 Contextualism and its discontents .................................................................... 65
    2.2.1 Contextualism ......................................................................................... 65
    2.2.2 Discontents .............................................................................................. 66
  2.3 Prospective contextualism .................................................................................. 70
  2.4 Performativity ........................................................................................................ 73
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.1 Performatives in natural language</td>
<td>75</td>
</tr>
<tr>
<td>2.4.2 Norms for modals: The perspective from performatives</td>
<td>76</td>
</tr>
<tr>
<td>2.5 Comparisons</td>
<td>80</td>
</tr>
<tr>
<td>2.6 Implementation</td>
<td>88</td>
</tr>
<tr>
<td>2.7 Conclusion</td>
<td>90</td>
</tr>
<tr>
<td>3 What ‘must’ adds</td>
<td>95</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>95</td>
</tr>
<tr>
<td>3.2 The data</td>
<td>97</td>
</tr>
<tr>
<td>3.2.1 Indirectness</td>
<td>98</td>
</tr>
<tr>
<td>3.2.2 Support</td>
<td>99</td>
</tr>
<tr>
<td>3.2.3 More cases</td>
<td>102</td>
</tr>
<tr>
<td>3.2.4 Experimental data</td>
<td>106</td>
</tr>
<tr>
<td>3.3 Explaining Indirectness via Support</td>
<td>110</td>
</tr>
<tr>
<td>3.3.1 Against strategies 1 and 2</td>
<td>111</td>
</tr>
<tr>
<td>3.3.2 From Support to Indirectness</td>
<td>118</td>
</tr>
<tr>
<td>3.3.3 Non-Redundancy</td>
<td>122</td>
</tr>
<tr>
<td>3.3.4 Predictions</td>
<td>125</td>
</tr>
<tr>
<td>3.4 Support</td>
<td>128</td>
</tr>
<tr>
<td>3.4.1 Extant proposals</td>
<td>128</td>
</tr>
<tr>
<td>3.4.2 Support as a manner implicature</td>
<td>131</td>
</tr>
<tr>
<td>3.5 Weakness</td>
<td>135</td>
</tr>
<tr>
<td>3.6 Embeddings</td>
<td>140</td>
</tr>
<tr>
<td>3.7 Conclusion</td>
<td>142</td>
</tr>
<tr>
<td>4 Modality and expressibility</td>
<td>145</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>145</td>
</tr>
<tr>
<td>4.2 Background</td>
<td>147</td>
</tr>
<tr>
<td>4.2.1 The relational theory</td>
<td>147</td>
</tr>
<tr>
<td>4.2.2 Yalcin (2007)'s challenge</td>
<td>149</td>
</tr>
<tr>
<td>4.2.3 Yalcin (2007)'s domain semantics</td>
<td>150</td>
</tr>
<tr>
<td>4.2.4 Ninan (2016)'s response</td>
<td>152</td>
</tr>
<tr>
<td>4.3 Relative potential expressibility</td>
<td>153</td>
</tr>
<tr>
<td>4.3.1 The domain semantics</td>
<td>157</td>
</tr>
<tr>
<td>4.3.2 The update and state-based semantics</td>
<td>159</td>
</tr>
<tr>
<td>4.3.3 Quantified modal languages</td>
<td>162</td>
</tr>
<tr>
<td>4.4 The upshot</td>
<td>164</td>
</tr>
<tr>
<td>4.5 Simple shifty systems</td>
<td>166</td>
</tr>
</tbody>
</table>
Introduction

Epistemic modals—words like ‘might’ and ‘must’ (and their analogues in other languages) on a broadly epistemic interpretation—play a crucial role in coordinating information in discourse, as in an exchange like (1):

(1)  a. [Ruth:] Where are the keys?
    b. [Jack:] They might be in the garage.
    c. [Ruth:] No, they can’t be; Julie had them last.
    d. [Jack:] Ok, then they must be in her room.

At the outset, neither Ruth nor Jack knows where the keys are. But each knows something about where they could be; and coordinating on this knowledge lets them figure out, in the end, where they are. Epistemic modals play a crucial role in this kind of coordination, which, in turn, plays a central role in cooperative reasoning and action. In this dissertation, I give a theory of what epistemic modals mean and how people use them to coordinate on their information.

I begin by exploring the meaning of epistemic modals. My guiding question is: to what degree does an epistemic modal claim like ‘It might be raining’ resemble an avowal of ignorance like ‘For all I know, it’s raining’? Progress on this question has been made by exploring differences in how constructions like these embed—in particular by exploring their behavior as part of larger
constructions like Wittgenstein (1953)'s ‘It might be raining and it's not' and Moore (1942)'s ‘It’s raining and I don’t know it', respectively. Based on the divergent behavior of these sentences in disjunctions, I argue that (contrary to the consensus in the literature) Wittgenstein sentences are classical contradictions. I develop a new semantics for epistemic modals to make sense of this which I call the bounded theory of epistemic modality. The bounded theory comprises a standard relational theory together with a constraint which predicts that local information is always taken into account in the evaluation of epistemic modals. I show that this theory accounts for the subtle behavior of embedded modals, and sheds new light, not only on the meaning of epistemic modals, but also on the way that local information is structured in natural language.

In the second chapter, I turn from the question of what epistemic modals mean to the question of how speakers use those meanings to coordinate on their common information. I show that—pace much recent work—we can make sense of this in a broadly relational semantic framework (like the one developed in the first chapter), together with a standard contextualist approach to the dynamics of conversation. I argue that epistemic modal assertions are performatives which make claims about what structural properties (like compatibility with $p$, or entailing $p$) the common ground will have after the claim is accepted or rejected. Thanks to the logic of the common ground, epistemic modal claims will have the effect of proposals to bring it about that the common ground has the property in question. This approach makes sense of what speakers (dis)agree about when they (dis)agree about modal claims, and what they are doing when they assert modal claims: negotiating what structural properties the common ground ought to have. And it does so within the simple and attractive contextualist framework.

The third chapter explores the role of ‘must' in conversation. Assertions of ‘Must $p$' and assertions of $p$ alone seem to have the same basic goal: namely, coming to agreement that $p$ is true. Yet their felicity conditions differ in subtle ways. I argue for a new characterization of those differences, and a broadly pragmatic account of them. My central claim is that an assertion of ‘Must $p$' is a proposal to accept $p$, but to do so based on a shared argument, rather than based on the speaker’s epistemic authority. I show that the requirement that the speaker’s evidence for $p$ be indirect follows pragmatically from this constraint. I argue that this approach accounts for the sense
in which 'must' is strong (it has a strong update effect), as well as the sense in which 'must' is weak (the speaker is proposing an update, not on the basis of her own epistemic authority, but rather on the basis of a shared argument).

In the final chapter, I take an abstract perspective on the choice of semantic theories for epistemic modals. I develop formal tools to compare the expressive power of different theories of epistemic modals with respect to the embedding data which they can account for. I show that the relational theory is strictly more expressive than several recent revisionary theories, including the domain, update, and state-based semantics. This fact necessitates a substantial shift in perspective on the debate regarding the semantics of epistemic modals. On the one hand, since essentially any data that one of these revisionary theories can make sense of can also be handled by the standard theory, it is very unlikely that there are data which will show the standard theory to be wrong and a revisionary theory correct. On the other hand, insofar as those revisionary theories are weaker than the standard theory vis-à-vis their expressive power, there should be a prejudice in their favor, as long as they can make sense of all the relevant data. Considerations from the first chapter, though, show that they may not, in the end, be up to that task; in the final part of the chapter I explore what the behavior of epistemic modals under attitude predicates tells us about the choice between frameworks.

The four chapters of the dissertation build on each other in a variety of ways. However, each chapter is written in such a way that it can be read on its own (which leads to some mild, but unavoidable, redundancies across the chapters).
Chapter 1

Bounded modality

I begin by exploring the meaning of epistemic modals. I present data which I argue make trouble for nearly every extant theory. I develop a new semantics to make sense of those data, which I call the bounded theory of epistemic modality. The bounded theory comprises a standard relational theory together with a constraint which predicts that local information is always taken into account in the evaluation of epistemic modals. I show that this theory accounts for the subtle behavior of embedded modals—including the new data I present, as well as known patterns under attitude predicates, conditionals, and quantifiers—and also sheds new light on the way that local information is structured in natural language.

1.1 Introduction

Moore (1942) observed that there is something wrong with asserting a sentence like (1):¹

(1) #It’s raining but I don’t know it.

Generally speaking, sentences with the form ‘p, but I don’t know p’ or ‘I don’t know that p, but

¹Moore’s sentences were formulated in terms of belief, but I will stick with knowledge here: this makes the parallel to Wittgenstein sentences closer, and it avoids issues involving neg-raising. The factivity of ‘knows’ does not play an essential role here: all my points can be made with a non-factive variant, provided that care is taken to avoid neg-raising readings. I will generally move freely between using ‘and’ and ‘but’ as a conjunction in this chapter, assuming they have relevantly similar semantics; readers can confirm for themselves that alternating between them does not change felicity judgments.
p\^7 (now known as \textit{Moore sentences}) tend to be unassertable.\footnote{Which is not to say unasserted. For (something very close to) a recent example: 'I don’t stand by anything. I just—you can take it the way you want. I think our side’s been proven very strongly' (Donald Trump, \url{http://www.politico.com/story/2017/05/01/trump-surveillance-claims-cbs-interview-237831}).}

In an early discussion of Moore sentences, Wittgenstein (1953, II.x.109) observed a similar phenomenon: the strangeness of sentences like (2):\footnote{Es dürfte regnen; aber es regnet nicht.' German speakers inform me that 'dürfte' has a slightly stronger meaning than that of 'might' (but note that 'should' is not an adequate translation, since 'It should be raining but it's not' can be felicitous, whereas Wittgenstein's sentence cannot). I will ignore the complexities of the German modal system here, resting content with 'might' as a translation.}

\begin{quote}
(2) #It might be raining but it's not raining.
\end{quote}

Call sentences like (2)—with the form 'Might p and not \(p\)^7, or \(\neg p\) and might not \(p\)^7—\textit{Wittgenstein sentences}.\footnote{Yalcin (2007) calls sentences with the form \(\neg p\) and might not \(p\) \textit{epistemic contradictions}. I will stay away from that name, for a few reasons: first, the order of conjuncts does not matter for my purposes; second, I do not want to prejudge the question of whether sentences like this are contradictions, and, if so, whether they are echt contradictions or just 'epistemic' ones.}

Moore and Wittgenstein sentences have played a central role in recent inquiry into the meaning of epistemic modals (words like 'might' and 'must', and their analogues in other languages, on a broadly epistemic interpretation). On the face of it, an epistemic modal construction like 'It might be raining' means roughly the same thing as 'We don't know it’s not raining', or, more colloquially, 'For all we know, it’s raining', where 'we' refers to some contextually salient group. A central question in contemporary inquiry into the meaning of epistemic modals is how close this parallel really is.

A broad range of work in recent years has shown striking differences between the ways Wittgenstein sentences and Moore sentences embed, and thus that this parallel has limits. Researchers have made a variety of proposals to account for these differences. Nearly all those proposals, however, agree on one point: namely, that the infelicity of an \textit{unembedded} Wittgenstein sentence like (2) is to be explained in at least roughly the same way as the infelicity of an \textit{unembedded} Moore sentence like (1): both are \textit{classically consistent}; their infelicity is due to the fact that it is in some sense incoherent to commit yourself to both conjuncts.

In this chapter, I will argue that this consensus is wrong. If this consensus were right, then
disjunctions of Moore sentences, and disjunctions of Wittgenstein sentences, would sound coherent. This prediction is borne out in the former case, but not the latter. This strongly suggests that \( \text{Might } p \) and not \( p \) is a classical contradiction, just as it appears to be. But this, in turn, creates a puzzle. If \( \text{Might } p \) and not \( p \) is a classical contradiction, then, assuming ‘and’ and ‘not’ have a classical semantics, it will follow that \( \text{Might } p \) entails \( p \); but that is clearly false.

I solve this puzzle by giving a new theory of the meaning of epistemic modals and their interaction with embedding operators. The basic idea is that epistemic modals are quantifiers over accessible worlds, as the standard theory has it; but, crucially, they can only quantify over worlds in their local contexts. Together with an appropriate theory of local contexts, I show that this approach predicts that Wittgenstein sentences are classical contradictions—thus making sense of the infelicity of disjoined Wittgenstein sentences—without also predicting that \( \text{Might } p \) entails \( p \). I argue that the resulting theory, furthermore, accounts for the subtle and surprising behavior of epistemic modals in a wide range of further embedding environments—suggesting that this approach gets to the heart of the question of what epistemic modals mean, and how the meaning of \( \text{Might } p \) diverges from that of \( \text{For all we know, } p \). Finally, I argue that, more broadly speaking, this theory sheds important new light on the dynamics of information in natural language.

My plan is as follows. In §1.2, I introduce the standard approaches to Moore and Wittgenstein sentences. In §2.2.2, I show that this class of approaches cannot make sense of the infelicity of disjunctions of Wittgenstein sentences. In §1.4, I introduce my theory of epistemic modals and their interaction with embedding operators, which I call the bounded theory. In §1.5, I show how this theory accounts for disjoined Wittgenstein sentences. In §1.6, I argue that the relationality of my theory is essential to my solution: contra much recent work which has used the embedding behavior of epistemic modals to argue against relational approaches to their semantics, it is only the relational framework which can make sense of epistemic modals’ distinctive dependence on the intrasentential dynamics of information. In §1.7, I show that my approach provides a principled explanation of the embedding behavior of epistemic modals in a wide range of further environments. In §1.8 I argue that my account compares favorably to the only other extant approaches that can predict the infelicity of disjoined Wittgenstein sentences. I conclude in §1.9.
1.2 The classical strategy

Suppose Ruth asserts the Moore sentence (1):

(1) #It’s raining but I don’t know it.

There is strong reason to think that the content of Ruth’s assertion is classically consistent: there are obviously circumstances in which (1) is true, namely those where it’s raining out, but Ruth doesn’t know it. (1) is nonetheless unassertable. It is fairly easy to see, in broad terms, why this should be so.\(^5\) Asserting something generally involves a commitment to knowing its content. If Ruth asserts (1), she expresses that she knows the following: that it’s raining and that she doesn’t know that it’s raining. Assuming knowledge distributes over conjuncts, Ruth thus expresses that she knows that it’s raining and that she knows that she doesn’t know that it’s raining. But such a state could never obtain: if she knows that she doesn’t know that it’s raining, then she doesn’t know that it’s raining. Call this style of approach to the infelicity of Moore sentences the classical strategy.

Now suppose Ruth asserts the Wittgenstein sentence (2):

(2) #It might be raining but it’s not raining.

How far should our explanation of the infelicity of a Wittgenstein sentence mirror our explanation of the infelicity of Moore sentences? As we’ll see in detail below, there are important differences between Wittgenstein and Moore sentences, in particular with respect to how they embed. But there is a simple and persuasive argument that our explanations for the infelicity of unembedded Moore and Wittgenstein sentences should be roughly the same. Just as for Moore sentences, there is a clear argument that Wittgenstein sentences are classically consistent.\(^6\) Suppose that ‘Might \(p\)’ and ‘Not \(p\)’ were inconsistent. Then it follows in classical logic that ‘Might \(p\)’ entails \(p\). But that can’t be right. ‘Might \(p\) and might not \(p\)’ is often true (as in ‘It might be raining and it might not be raining—I have no idea!’). But if ‘Might \(p\)’ entailed \(p\), then ‘Might \(p\) and might not \(p\)’ would be inconsistent, since it would entail \(p\) and not \(p\). Now, if two sentences are jointly consistent,\(^5\) The details of how this goes do not matter for our purposes. For a classic formulation, see Hintikka 1962.
\(^6\) See Yalcın 2007 for a clear statement of this argument.
it follows from classical logic that their conjunction is also consistent. And that, in turn, suggests that Wittgenstein sentences are classically consistent, and thus that the situation for unembedded Wittgenstein sentences is much like that for unembedded Moore sentences: they are consistent but still unassertable.

And indeed, it looks like we can explain this fact using roughly the same classical strategy that is used to explain the infelicity of unembedded Moore sentences. The general idea is that ‘It might be raining’ amounts to a proposal to leave open the possibility that it’s raining; while ‘It’s not raining’ amounts to a proposal to rule this possibility out. Asserting their conjunction will amount to making both of these proposals simultaneously. And there is something wrong with this, since both proposals cannot be simultaneously carried out. One way of spelling out this idea is to maintain that ‘It might be raining’ is generally only assertable when the speaker’s knowledge is consistent with the proposition that it’s raining; whereas, since one generally needs to know what one asserts, ‘It’s not raining’ is only assertable when the speaker’s knowledge is inconsistent with the proposition that it’s raining. Assuming that a conjunction is only assertable when each conjunct is assertable, it follows that a Wittgenstein conjunction will never be assertable—since one’s knowledge cannot be both consistent with, and inconsistent with, the proposition that it is raining.

As for Moore sentences, there are many different ways of spelling out the details of this classical strategy. Again, these differences don’t matter for our purposes. Some approaches, like Veltman (1996)’s or Yalcin (2007)’s, spell out this strategy by appeal to a broadly dynamic or informational notion of entailment, a notion relative to which Wittgenstein sentences are indeed inconsistent. Crucially, though, on those approaches—as on nearly every other approach (with the exception of the handful which I discuss in §1.8)—Wittgenstein sentences are still predicted to be classically consistent (in a way which can be made precise in those frameworks), a fact which, as we will see momentarily, leads to trouble.7

7In Veltman’s update framework, we can say that p is classically consistent just in case there is a context c such that c[p] is non-empty; as we will see, Wittgenstein sentences of the form ‘It might p and not p’ will be classically consistent in the update framework. In Yalcin’s domain framework, we can say that p is classically consistent just in case p is true at some point of evaluation; as we will see, Wittgenstein sentences with either order will be consistent in this sense in the domain framework.
1.3 Disjunctions

The classical strategy to Moore and Wittgenstein sentences yields a striking, and heretofore unobserved, prediction. On the classical strategy, a Moore or Wittgenstein sentence is infelicitous only because no speaker can coherently be committed to both of its conjuncts. Suppose, however, that we take a disjunction of two Moore sentences, or two Wittgenstein sentences. In asserting a disjunction, the speaker expresses that she knows the disjunction, but not that she knows any one disjunct on its own—and thus does not express that she knows any one incoherent conjunction. The classical strategy thus predicts that disjoined Moore or Wittgenstein sentences will be felicitous.8

Is this prediction borne out? Consider first Moore sentences. Suppose I’m party to a court case which hasn’t been decided. I don’t know whether I’ll win or lose (and I know that I don’t know this). In this situation, the Moore sentences (3) and (4) are, as always, unassertable:

(3)  a. #I’ll win but I don’t know it.
    b. #p and I don’t know p.

(4)  a. #I’ll lose but I don’t know it.
    b. #q and I don’t know q.

But now consider the disjunction of (3) and (4):

(5)  a. Either I’ll win but I don’t know it, or I’ll lose but I don’t know it.
    b. p and I don’t know p, or q and I don’t know q.

Exactly as the classical strategy predicts, (5) is impeccable: disjoining Moore sentences bleaches them of their infelicity.9

8I know of no discussion of cases with the form of those discussed here, but Dorr and Hawthorne 2013 discusses closely related cases, namely sentences with the form "p and I don’t know that p, or q" and "p and might not p, or q" (as well as order variants), and thus is an important precedent for this discussion (thanks to Shane Steinert-Threlkeld for pointing this out to me). These cases illustrate the same basic point I make in this section, albeit in a different way: as Dorr and Hawthorne discuss, the issue with sentences like "p and might not p, or q" is that the first disjunct is somehow felt to be redundant, not that the whole disjunction is somehow felt to be something like a contradiction, as in the case of the Wittgenstein disjunctions I will discuss.

9Note that, while the fact that the alternatives in the disjunction (winning and losing) are exclusive makes the disjunction easier to process (something pointed out to me by Fabrizio Cariani), this is inessential for my point, as examples (11)
In more detail, the reason the classical strategy predicts that an assertion of a sentence like (3-a) is infelicitous is that the speaker is committing herself to knowing two things: first, that she will win; second, that she doesn’t know that she will win. The speaker could know either of these things on its own, but she could not know both together, since for her to know the second entails that she doesn’t know the first. But now consider the disjunction in (5-a). Each of its disjuncts, again, is such that no speaker could know it to obtain. But a speaker could know that either the first disjunct is true or that the second is true, since to know a disjunction does not require knowing one disjunct. And so when a speaker asserts (5-a), thus expressing that she knows it, she does not commit herself to any kind of inconsistency. On the contrary, for the speaker to know that this disjunction obtains is a fairly straightforward matter: it essentially just amounts to knowing that she is ignorant of the outcome of the court case.

The felicity of disjunctions like (5-a) offers a striking confirmation of the classical approach to Moore sentences. But matters are different when we turn to Wittgenstein sentences. In the same scenario, the Wittgenstein sentences (6) and (7) are—again, unsurprisingly—unassertable:

(6)  
   a. #I might win but I won’t.  
   b. #Might p and not p.

(7)  
   a. #I might lose but I won’t.  
   b. #Might q and not q.

But note that their disjunction, in (8), remains just as infelicitous:

(8)  
   a. #Either I might win, but I won’t, or I might lose, but I won’t.  
   b. #Might p and not p, or might q and not q.

This is in striking contrast to Moore sentences.

This pattern is robust across permutations in order as well as variations in our case. Consider first sentences which permute the order of the conjuncts in (5-a) and (8):

and (12) below show (it is possible that both Mary and Sue are my TAs).
(9) a. Either I don’t know I’ll win, but I will win; or I don’t know I’ll lose, but I will lose.
    b. I don’t know p, and p, or I don’t know q, and q.

(10) a. #Either I won’t win, but I might; or I won’t lose, but I might.
    b. #Not p and might p, or not q and might q.

Again, the disjunction of Moore sentences is felicitous, but the disjunction of Wittgenstein sentences is not. Thus these judgments seem to remain the same across order permutations.

To get more cases on the table, suppose that either Sue or Mary is my TA, but I can’t remember which is (and possibly both are). Compare (11) and (12):

(11) a. Either Sue is my TA but I don’t know it, or Mary is my TA but I don’t know it.
    b. Either I don’t know it, but Sue is my TA; or I don’t know it, but Mary is my TA.

(12) a. #Either Sue might be my TA but she isn’t, or Mary might be my TA but she isn’t.
    b. #Either Sue isn’t my TA but she might be, or Mary isn’t my TA but she might be.

Once again, disjunctions of Moore sentences (of either order) are felicitous, but disjunctions of Wittgenstein sentences (of either order) are not. Importantly, note that this case uses present tense, stative predicates in the complements of ‘might’, which selects for an epistemic reading of ‘might’, ruling out an irrelevant metaphysical reading.\(^\text{10}\)

For a final case, using past predicates, suppose that one of my students, either Mary or Sue, was sick last week, but I can’t remember which. Then I can say (13):

(13) a. Either Sue was sick but I don’t know it, or Mary was sick but I don’t know it.
    b. Either I don’t know it, but Sue was sick; or I don’t know it, but Mary was sick.

By contrast, (15) sounds quite odd, at least on its (default) epistemic reading:

(14) a. #Either Sue might have been sick but she wasn’t; or Mary might have been sick but she wasn’t.
    
\(^{10}\text{Cf. Moore 1959.}\)
b. #Either Sue wasn’t sick but she might have been; or Mary wasn’t sick but she might have been.

While there is a felicitous reading of (14), it is clearly a metaphysical reading which communicates something quite different from (13) (it communicates that either it was the case that Sue wasn’t sick, but there was a close non-actual situation where she was sick; or else that the same held of Mary).

The pattern is question thus looks robust: disjoined Moore sentences are generally perfectly felicitous, while disjoined Wittgenstein sentences (on the intended epistemic reading of the modals) are generally marked, and, importantly, much less felicitous than the corresponding disjoined Moore sentences.\(^{11}\)

The fact that disjoined Wittgenstein sentences (call them Wittgenstein disjunctions) are infelicitous presents a serious challenge to the classical strategy for explaining their infelicity sketched in the last section. The problem is that the classical strategy predicts that the infelicity of Wittgenstein sentences is due to the incoherence of simultaneously committing to two conjuncts which are, however, jointly consistent in a classical sense. If this were right, then, just as for Moore sentences, the infelicity of Wittgenstein disjunctions should wash out when they are disjoined, since when you assert a disjunction, you are not committed to any disjunct: again, you can know a disjunction without knowing a disjunct. As we have seen, this is exactly the right prediction when it comes to Moore sentences. But it is, apparently, not the right prediction when it comes to Wittgenstein sentences.

Let me emphasize the significance of this fact. It is already known that Wittgenstein sentences embed differently from Moore sentences. But nearly every theory of epistemic modals—including those which have been motivated by various embedding data\(^ {12}\)—nevertheless remains committed to the classical consistency of Wittgenstein sentences, and thus to a broadly classical approach to explaining their infelicity. And, again, this commitment is no accident; there is a persuasive argument that Wittgenstein sentences are classically consistent. But the infelicity of Wittgenstein

\(^{11}\)Intuitions in some of these cases will certainly be graded—intuitions about natural language are rarely binary; but every informant I have discussed these data with reports a very strong contrast between Wittgenstein disjunctions and Moore disjunctions, and it is that contrast which is a puzzle for standard theories.

\(^{12}\)Like the update semantics of Veltman 1996; Groenendijk et al. 1996; Beaver 1992, 2001; Aloni 2001; Yalcin 2012b, 2015; Willer 2013; Ninan 2017, or the domain semantics of MacFarlane 2011; Yalcin 2007; Klinedinst and Rothschild 2012. Also vulnerable to my objection are e.g. the proposals of e.g. Egan et al. 2005; Hacquard 2006; Stephenson 2007b,a; Lasersohn 2009; Ninan 2010; Swanson 2015; Ninan 2010, 2016; Stejnic 2016.
disjunctions makes this strategy, and thus these semantic theories, look untenable.

In the rest of this section, I will spell out these considerations more carefully by introducing three semantic theories—the relational, domain, and update semantics—and using them to illustrate the problem (readers uninterested in these details may wish to skip ahead to the next section). I choose these theories because they are representative of three broad approaches to the semantics of epistemic modals, but I emphasize that the problem is a problem not just for these theories, but for *every* extant theory of epistemic modals, with the exception of those I discuss in §1.8.

On the relational semantics—due in particular to Kratzer (1977, 1981), building on earlier work in modal logic—epistemic modals denote quantifiers over a set of worlds, generally assumed to represent a relevant epistemic state or body of evidence. The set of worlds is supplied by a variable which, in turn, is assigned to a function from worlds to sets of worlds (a modal base or accessibility relation) by a relevant variable assignment $g$:

$$[\text{Might}_i p]^{g,w} = 1 \iff \exists w' \in g(i)(w) : [p]^{g,w'} = 1$$

Informally: "$\text{Might}_i p$" says that $p$ is true at some epistemically accessible world.

Given this gloss, it should already be clear why this approach will fail to predict the infelicity of disjoined Wittgenstein sentences. On this approach, "$\text{Might}_i p$" means roughly: "$\text{For all we [the relevant agents] know, } p$". But then the meaning of Wittgenstein sentences and Moore sentences will be essentially the same, and so they will be predicted to behave in the same way in disjunctions.

To spell this out more formally, we need a semantics for connectives; I will assume, as is standard in this framework, a simple Boolean semantics for conjunction, disjunction, and negation:

---

13"Accessibility relation" is somewhat sloppy, of course—an accessibility relation is officially a binary relation between worlds—but this makes no difference, and it will be simpler for us to work with the corresponding function from worlds to sets of worlds. I am following von Fintel (1994)'s implementation here, which will lend itself well to the positive view I sketch below. Although I will talk about a 'contextually relevant variable assignment', how exactly $g$ is provided—whether by the context of assertion, as contextualists maintain, or of assessment, as relativists maintain, or not at all, as expressivists maintain, will not matter for present purposes. The present presentation is a simplification of Kratzer's theory, on which modals are evaluated relative to a second parameter, an ordering source; and on which an accessibility relation is a function from worlds to sets of propositions, not sets of worlds. Those complications will not matter for present purposes, so I will ignore them. I use roman letters to stand for sentences and italic letters to stand for the corresponding propositions (which for simplicity I assume are functions from worlds to truth values, or equivalently sets of worlds), implicitly relativized to an index of evaluation and a context, i.e. where $\langle m_1, m_2, \ldots, m_k, w \rangle$ is an index of evaluation and $c$ is a context, $\text{true}_w[p]^{c,m_1,m_2,\ldots,m_k,w'}$. Where $p$ is a proposition or property, $\overline{p}$ is its complement. I will generally leave off context parameters in what follows for readability.
Now consider a Wittgenstein disjunction with the form \( \neg \text{Might}_i p \) and not \( p \) or mighty \( q \) and not \( q \) (to map this onto our example from above, let \( p = 'I \text{ win}' \) and \( q = 'I \text{ lose}' \)). Consider a context which contains two worlds, \( w \) and \( w' \). At \( w \), \( p \) is true and \( q \) false.\(^{14}\) At \( w' \), \( p \) is false and \( q \) true. Assume that the contextually relevant variable assignment \( g \) takes \( i \) and \( j \) to the accessibility relation \( f \) which takes each world in the context to the context set itself, i.e. \( g(i)(w) = g(i)(w') = g(j)(w) = g(j)(w') = \{w, w'\} \). Then our schematic Wittgenstein disjunction will be true at \( w \), thanks to the truth at \( w \) of the second disjunct, \( \neg \text{Might}_j q \) and not \( q \). This is true at \( w \) since \( w \) can access a \( q \)-world under \( g(j) \) (namely \( w' \)), and since \( q \) is false at \( w \). And our Wittgenstein disjunction will be true at \( w' \), since, by parallel reasoning, \( w' \) makes the first disjunct true. This shows that our Wittgenstein disjunction is consistent, since it is true at some points of evaluation in some context, relative to some variable assignments.

Note moreover that our Wittgenstein disjunction is true \textit{throughout} this context. If we take a context to represent the worlds compatible with what is commonly accepted in the conversation,\(^{15}\) this shows that Wittgenstein disjunctions are predicted to be not only consistent, but (unlike a Moore sentence) can be commonly accepted in some contexts. And I can see nothing untoward about the context as described, or about the way we selected the accessibility relation—it seems perfectly reasonable to think that the accessibility relation will at least sometimes track the context itself.\(^{16}\)

This shows that the relational framework predicts that Wittgenstein disjunctions are classically consistent, and, worse, coherent in a broader, pragmatic sense. There might, of course, still be a broadly pragmatic explanation of their infelicity within the relational framework. But it is hard to see how this kind of a story would go. Any pragmatic response to this problem would have to

\(^{14}\)I assume for simplicity that \( p \) and \( q \) are non-modal sentences, so their truth value depends only on the world of evaluation.

\(^{15}\)As in Stalnaker 1974; more on this below.

\(^{16}\)See e.g. Stalnaker 2014, and Chapter 2, for views along these lines.
acknowledge that there are some theoretically possible accessibility relations which render Wittgenstein disjunctions consistent, but maintain that such accessibility relations are never in fact chosen. In the end I will defend a view which has this result, but which gets this result from richer semantic resources than the relational theory provides on its own. On the basis of the relational theory alone, plus general facts about rationality, I cannot see how we would get this result. The question, again, is why we would not simply choose an accessibility relation which allows us to interpret the speaker as saying something coherent—that is, as saying essentially what Moore disjunctions say. In other words, why don't we interpret Wittgenstein disjunctions just as (somewhat periphrastic) expressions of ignorance about the relevant facts? I do not see how the relational semantics can explain this fact.

It may not be all that surprising that the relational theory cannot make sense of Wittgenstein disjunctions; after all, as we will see in §1.7 below, it is already well known that the relational theory runs into trouble in a variety of ways when it comes to embedded modals. Somewhat more surprising, however, is that this problem extends to the domain and update semantics, both of which have been motivated by their ability to account for the embedding behavior of epistemic modals.

At a high level, we have already seen why these accounts run into trouble: both theories predict that at least some Wittgenstein sentences are classically consistent, and thus that their disjunctions will be coherent. Thus, first, in the domain semantics of Yalcı 2007; MacFarlane 2011, epistemic modals denote quantifiers over a set of worlds which is supplied, not by an accessibility relation, but rather as a world-independent parameter of the index:

(19) $\text{[Might } p\text{]}^{s,w} = 1 \text{ iff } \exists w' \in s : \text{[p]}^{s,w'} = 1$

This semantics is very close to the relational semantics, but with a set of worlds substituted where the relational semantics has a function from worlds to sets of worlds. Domain semantics is standardly coupled with Boolean semantics for the connectives, generalized to the domain framework as follows:

$\text{[Might } p\text{]}^{s,w} = 1 \text{ iff } \exists w' \in s : \text{[p]}^{s,w'} = 1$

This is a challenge for the kind of approach to embedded modals given in Dorr and Hawthorne 2013 (and the kind of response that I imagine other defenders of the relational theory, like Dowell 2011, might give). Dorr and Hawthorne address this worry, but in a way which does not seem to me to fully capture the contrast between Moore and Wittgenstein disjunctions, especially given that the infelicity of Wittgenstein disjunctions is invariant across order permutations; for reasons of space, I cannot do justice to their discussion here.

Klinedinst and Rothschild 2012 diverges from this assumption in a way I briefly discuss in §1.6.
In the domain semantics, an assertion of $p$ is a proposal to make the context accept $p$, where a set of worlds $s$ accepts a sentence $p$ just in case $\forall w \in s : [p]^{s,w} = 1$. This gives us a precise notion of pragmatic coherence: we can say that a sentence is coherent just in case it is accepted by a non-empty context. Consider the same context from above, call it $s$. Recall that $s = \{w, w'\}$; at $w$, $p$ is true and $q$ false, and at $w'$, $p$ is false and $q$ true. Then $[\text{Might } p \text{ and not } p, \text{ or might } q \text{ and not } q]^{s,w} = 1$, since $[\text{Might } q \text{ and not } q]^{s,w} = 1$, since $[\text{Might } q]^{s,w} = 1$ (since $s$ contains $q$-worlds) and $[\text{Not } q]^{s,w} = 1$ (since $q$ is false at $w$). By parallel reasoning on the other disjunct, $[\text{Might } p \text{ and not } p, \text{ or might } q \text{ and not } q]^{s,w} = 1$. And thus $s$ accepts $[\text{Might } p \text{ and not } p, \text{ or might } q \text{ and not } q]$. (Note that the order of the conjuncts doesn’t matter here; $s$ also accepts $[\text{Not } p \text{ and might } q, \text{ or not } q \text{ and might } q]$). Thus in the domain framework, Wittgenstein disjunctions are classically consistent, since they are true at some points of evaluation. They are also pragmatically coherent (or, in Yalcin (2007)'s terminology, informationally consistent), since they are accepted at some non-empty contexts.

Let’s turn, finally, to the update semantics for epistemic modals, due to Veltman (1996). That theory is given within the dynamic framework of Heim (1982, 1983). In Heim’s framework, the semantic value of a sentence $p$ is a context change potential (CCP), denoted $[p]$: a function from contexts to contexts. The denotations of atomic sentences, negation, and conjunction are defined as follows in Heim (1983)'s system (using standard postfix notation: $c[p]$ is $[p]$ applied to $c$).19 For any context $c$:

(23) For atomic $p$, $c[p] = \{w \in c : p \text{ is true in } w\}$

(24) $c[\text{Not } p] = c \setminus c[p]$

Veltman (1996) assumes a different entry for conjunction, treating $c[p \text{ and } q]$ as $c[p] \cap c[q]$. Here I follow most of the subsequent literature (e.g. Groenendijk et al. (1996)) in using Heim’s entry for conjunction, which is more interesting for our purposes. Readers can verify for themselves that substituting in Veltman’s entry does not help, and indeed deprives update semantics of its limited victory in predicting the infelicity of Wittgenstein disjunctions with right-embedded modals.
We augment Heim's semantics with any dynamic entry for disjunction—it won't matter which entry we adopt, so I will assume the simplest possible entry, from Veltman (1996).

\[(25) \quad c[p \text{ and } q] = c[p][q] \]

Finally, Veltman proposes the following semantics for 'might':

\[(27) \quad c[\text{Might } p] = \{w \in c : c[p] \neq \emptyset\} \]

On this semantics, \(\text{Might } p\) is a test on the context: it leaves the context unchanged if it is compatible with \(p\), and otherwise induces a 'crash' to the empty set. In a dynamic framework, the question of whether a sentence \(p\) is pragmatically coherent is the question of whether there is a non-empty context which remains unchanged when updated with \(p\): that is, whether there is a non-empty context \(c\) such that \(c[p] = c\) (if there is, say, again, that \(c\) accepts \(p\)). It turns out that the same context \(s\) from above accepts \(p\). Recall again that \(s = \{w, w'\}\), with \(p\) true and \(q\) false at \(w\), and \(p\) false and \(q\) true at \(w'\). Then \(s\) accepts \(\text{Might } p\) and not \(p\), or might \(q\) and not \(q\). Informally, we first update \(s\) with \(\text{Might } p\), which leaves \(s\) unchanged since it contains a \(p\)-world. Then we update it with \(\text{Not } p\), removing the \(p\)-world, so we have just \(\{w'\}\). Simultaneously, we update \(s\) with \(\text{Might } q\), which leaves \(s\) unchanged, since it contains a \(q\)-world. Then we update it with \(\text{Not } q\), removing the \(q\) world, so we have just \(\{w\}\). Then our entry for disjunction leads us to take the union of \(\{w\}\) with \(\{w'\}\), giving us \(s\) back again. Thus, once more, the update approach predicts that Wittgenstein sentences are consistent (in the dynamic sense that they do not take every context to the empty set), and, moreover, that they are accepted by some non-empty contexts, and thus are in some sense pragmatically coherent (or dynamically consistent in any of the senses given in Veltman 1996).

One interesting feature of the update semantics is that—unlike for the relational or domain semantics—order matters: on the update semantics, no non-empty context accepts a Wittgenstein disjunction with right-embedded modals, like \(\text{p and might not } p\), or \(q\) and might not \(q\). At a high
level, this is because, in this framework, $p$ is taken into account when we process "Might not $p$". This is a success, and this kind of idea will influence my own proposal below. But it is a limited success. As we saw above, the infelicity of Wittgenstein disjunctions is invariant across order permutations; but in the update framework, it is only left conjuncts which influence the interpretation of modals in right conjuncts, and not vice versa. So this cannot be the end of the story.

In short, the relational, domain, and update semantics all predict that at least some Wittgenstein disjunctions will have roughly the same update effect as a Moore disjunction—making sure that the context contains both $p$- and $q$-worlds, and that it also contains $\bar{p}$- and $\bar{q}$-worlds—and thus will be perfectly coherent.\footnote{The kind of move that we have made here—disjoining Wittgenstein sentences to formulate an objection to theories that predict them to be classically consistent—can be replicated in other environments. Thus for instance "Might (might $p$ and not $p$)" is predicted to be consistent and acceptable in the relational, domain and update frameworks. Just as for Wittgenstein disjunctions, this seems to be the wrong prediction.}

### 1.4 Bounded modality

The classical approach to Wittgenstein sentences, on which Wittgenstein sentences are classically consistent, and incoherent in some other sense, thus yields a clear prediction that is not borne out: namely, that embedding Wittgenstein sentences in an environment like disjunction should yield a perfectly felicitous result. Recall, however, that it is not mere happenstance that nearly every theory of epistemic modals has converged on this broad strategy for accounting for Wittgenstein sentences (albeit in very different ways). Rather, these theories are motivated by the observation that "Might $p$" is classically consistent with "Not $p$" (otherwise, recall, "Might $p$" entails $p$, an untenable situation). How can we respect this observation, while also explaining the infelicity of Wittgenstein disjunctions, and their differences from Moore disjunctions?

To answer this challenge, I will give a new theory of epistemic modals, connectives in natural language, and their interaction. This theory will predict that, while "Might $p$" and "Not $p$" will often both be true at some context worlds, relative to a given accessibility relation, we are forced to evaluate their conjunction relative to an accessibility relation which ensures that their conjunction is never true at any world. The disjunction of two sentences like this will also fail to be true at any
world, and so this explanation of the infelicity of Wittgenstein sentences immediately extends to explain the infelicity of Wittgenstein disjunctions.

My theory, which I call the *bounded theory* of epistemic modality, comprises two ideas. The first concerns the architecture of embedding operators in natural language. Following literature on presupposition projection, I propose that embedding operators in natural language systematically make available a certain quantity of information (a *local context*) for the interpretation of embedded material. The second part of the theory concerns the meaning of epistemic modals. I propose that epistemic modals carry a semantic constraint which ensures that their local contexts limits their domain of quantification, so that modals only quantify over their local contexts. In the rest of this section, I will flesh out these two ideas. I will begin by giving a brief intuitive characterization of local contexts; spell out how my theory of epistemic modals will exploit local contexts; and then spell out a more precise algorithm for calculating local contexts.

Before doing so, let me note that my theory is a theory about the *semantics* of epistemic modals and embedding operators. My theory will remain agnostic about the pragmatics or 'post-semantics' (in MacFarlane (2014)'s terminology) of epistemic modals, and thus is compatible with a variety of answers to that question, including contextualist, relativist, and expressivist answers. It may turn out that my semantics lends itself better to one of these post-semantic theories than another, but whether that is so—and thus what indirect bearing my proposal has on the recent controversy about the post-semantics of epistemic modals—is not something I will explore in this chapter.

### 1.4.1 Local contexts: The basics

It is standard in theorizing about natural language to posit that the interpretation of an assertion depends on the assertion's *context*. There are different ways to spell out precisely what a context amounts to. For our purposes, the most useful notion treats a context as a set of possible worlds which in some sense represent the common commitments of the conversation.\(^{22}\)

Starting in the 1970s, a variety of theorists proposed that there is some sense in which contexts

\(^{21}\) That is, my theory is about the *semantic* content of epistemic modals, not its *assertoric content*, in the terminology of e.g. Ninan (2010).

\(^{22}\) The details of how this goes don't matter for us, but see Stalnaker 1974, 2002, 2014 for standard characterizations, and a comparison of competing approaches to contexts.
can shift within a sentence. The motivation for this was the phenomenon of presupposition projection. ‘Susie stopped smoking’ presupposes that she used to smoke. Now consider (28) and (29), which embed this sentence in different ways:

(28) If Susie used to smoke, then she stopped smoking.

(29) If Susie started exercising, then she stopped smoking.

Note that (29) still presupposes that Susie used to smoke (the presupposition projects to (29)), whereas (28) does not. A number of theories proposed to make sense of this roughly as follows.\textsuperscript{23} Presuppositions have to be entailed by their contexts. When presupposition “triggers” are unembedded, that means the presupposition in question has to be entailed by the conversation’s context (the global context). When they are embedded, however, this requirement is interpreted relative to a local context. In (28), the local context for the consequent entails that Susie used to smoke. Thus this presupposition is entailed by its local context, and so the sentence as a whole puts no further constraints on the global context. But things are otherwise in (29), where the local context for the consequent conjunct does not entail that Susie used to smoke, and thus the presupposition of the consequent still puts constraints on the global context.

This suffices to give a sense of the original motivation for countenancing local contexts. In a moment we will return to the question of what exactly they amount to; what is important for now is simply that a local context is a set of possible worlds, derived in a systematic way from the global context plus the relevant linguistic environment. Let’s turn now to see how my theory of epistemic modals exploits this parameter.

\subsection*{1.4.2 Local relational semantics}

My theory builds on the core truth conditions of the standard relational theory. But it augments those truth conditions with a \textit{locality constraint} which says that local contexts rule out certain possible accessibility relations. In particular, the constraint limits admissible accessibility relations to those which quantify over a subset of the local context: i.e. to those which ensure that only local context

\textsuperscript{23} Especially Stalnaker 1974; Karttunen 1974; Heim 1982, 1983.
worlds are accessible.

Formally, representing local contexts with $\kappa$, our semantics runs:\(^{24}\)

\[(30)\]  

\[\text{[Might, p]}_{g,\kappa,w}^\rho\]

a. defined only if $\forall w' : g(i)(w') \subseteq \kappa$; \hspace{1cm} [\textit{locality constraint}]

b. if defined, true iff $\exists w' \in g(i)(w) : [p]^\rho,\kappa,w' = 1$. \hspace{1cm} [\textit{relational truth conditions}]

(30-b) encodes the ordinary relational truth conditions for epistemic modals. The innovation is the locality constraint in (30-a). The locality constraint, again, ensures that an epistemic modal claim is only well-defined if the context's variable assignment associates the modal with an accessibility relation which only quantifies over worlds in the local context. In other words, the locality constraint ensures that accessibility relations can never reach across the boundaries set by local contexts.\(^{25}\)

The locality constraint is broadly inspired by the dynamic update semantics: the locality constraint ties the interpretation of epistemic modals to their local informational environment.\(^{26}\) But the present approach ends up being very different from the update approach in its empirical reach. In particular, adopting the present approach makes it possible to adopt a \textit{symmetric} theory of local contexts—something which, as I discuss in §1.6, turns out to be impossible to do satisfactorily in the dynamic framework, and which turns out to be essential for capturing the behavior of Wittgenstein disjunctions.

What kind of definedness condition does the locality constraint encode? We do not need a very precise answer to this question for present purposes. A natural approach is to understand it broadly along the lines on which semanticists analyze gender, person, and number features on

---

\(^{24}\) 'Must' will be defined as the dual of 'might', as usual. Note that the locality constraint will project, so that the semantics will run as follows: $[\text{Must, p}]^\rho,\kappa,w$ (a) defined only if $\forall w' \in \kappa : g(i)(w') \subseteq \kappa$; (b) if defined, true iff $\forall w' \in g(i)(w) : [p]^\rho,\kappa,w' = 1$. The locality constraint does not have very striking ramifications for the meaning of 'must'. I will continue to focus on 'might' here; careful exploration of 'must' will have to await future work.

\(^{25}\) The locality constraint is closely related to Stalnaker (1975)'s proposal that the selection function for indicative conditionals always selects a world in the context (thanks to Irene Heim for drawing my attention to this connection). I think there is an intimate relation between these constraints, one which I hope to draw out in future work on the bearing of the proposal here on the semantics of the indicative conditional.

\(^{26}\) See especially Beaver 2001, who brings out this feature of the update semantics; as well as Kliment and Rothschild 2012; Stalnaker 2014 for related discussions in the update and dynamic semantics. An early precedent for this kind of approach comes in McGee 1985; McGee gives a semantics for conditionals (not modals) on which their interpretation depends on a body of information which is shifted by the antecedents of conditionals. See Kratzer 1986, 1991; von Fintel 1994 for similar proposals about conditionals, cast explicitly in terms of epistemic modals.
pronouns—features which, like the locality constraint, are standardly treated as definedness conditions on variable assignments.\textsuperscript{27} But all that matters for us is that this definedness condition has the effect of ensuring that modals will be associated with accessibility relations which respect the locality constraint.

Call the semantics for epistemic modals in (30) the \textit{local relational semantics}. I will use \textit{bounded theory} to refer to the theory which results from combining this semantics with the specific approach to local contexts which I turn to now.

\subsection*{1.4.3 Local contexts, more precisely}

There has been much controversy over what local contexts amount to in the literature on presupposition. The controversy is two-fold. First, what are the empirical facts? Second, can we formulate a theory which predicts those facts in a systematic and explanatory way?\textsuperscript{28} Here I will follow (one variant of) Schlenker (2009)'s answer to these questions. The payoff of this choice will be in the empirical coverage it affords, and so I will not say much now to justify it.

Schlenker, building on Stalnaker 1974, proposes that the content of a local context at a given point in a sentence is whatever information would be felt to be \textit{redundant} at that point: that is, whatever information you could add to that sentence while being guaranteed not to change its truth conditions, given what is already taken for granted in the context. This provides a heuristic for determining what information is \textit{already available} at a given point in the sentence, given a classical semantic entry for the relevant embedding operator; then we treat that information as the local context introduced by the embedding operator.

There are two ways to flesh this idea out, depending how we think about redundancy. Is redundancy a \textit{temporal} matter—a matter of what material is redundant \textit{given how things stand at a particular moment in time} in processing a sentence? If so, then we should take an \textit{asymmetric}
approach, and say that a local context for an expression is whatever is redundant at that point in
the sentence, given everything to the left of that point. Or is redundancy a static, informational
matter—a matter of what material is redundant, given the rest of the information in a sentence? If
so, then we should take a symmetric approach, and say that the local context for an expression is
whatever is redundant at that point in the sentence, given everything else in the sentence. It is the
second, symmetric notion that will be useful for present purposes, for reasons that may already be
clear, given the symmetric nature of the judgments involving Wittgenstein disjunctions; I discuss
this choice further in §1.6 below.

Schlenker makes this symmetric approach to local contexts precise as follows: 29

Definition 1.4.1. Symmetric Local Contexts:
The symmetric local context of expression \( e \) in syntactic environment \( a\_b \) and context \( c \) is the
strongest \( [y] \) such that for all \( d \) of the same type as \( e \): \( a(d \land y)b \leftrightarrow_c adb \).

This algorithm simply makes precise the intuition that a local context is whatever content adds
no new information at a given point in a sentence, given everything else in that sentence and the
background assumptions in the context. To see how it works, let's go through an example that will
be crucial in what follows. Assume as a starting point the classical Boolean semantics for 'and'
given above. Then ask: what is the local context for a left conjunct? In other words, what is the
local context for \( q \) in some sentence \( \gamma q \land p \land c \) and context \( c \)? We ignore \( q \), and look at the schema
\( \gamma \land \) and \( p \land c \). Note that, whatever ends up going in for \( \gamma \), this sentence is guaranteed to be equivalent
to \( \gamma(p \land c) \land \gamma(c) \) and \( p \land c \). In other words, given the rest of the information in the sentence, plus
the background information \( c \), adding \( \gamma p \) and \( c \) to the left conjunct is guaranteed not to change
the truth-value of the sentence, no matter what the left conjunct amounts to. It is easy to show that

\(^{29} \land \) here denotes generalized conjunction, which is ordinary conjunction when \( x \) and \( y \) have propositional type, and
otherwise is conjunction built up in the obvious way from ordinary conjunction (we assume that \( x \) and \( y \) have types
that 'end in t'). Thus for instance 'blue and tall' is the generalized conjunction of 'blue' with 'tall', and has as its
extension a function from individuals to true just in case they are both blue and tall. For more formal details on this
point and others regarding this framework, see Schlenker 2009. My presentation is a slight variation on Schlenker's
symmetric framework, which stipulates that, in the symmetric algorithm, any presupposed material in \( a \) and \( b \) is ignored.
This won't affect present discussion, so I will adopt this simpler version here. \( \leftrightarrow_c \) is material equivalence modulo the
context \( c \). Since the semantics of epistemic modals will ultimately depend on local contexts, to avoid circularity we
should restrict the domain of the quantifier in this algorithm to non-modal sentences. This formulation is somewhat
slippery, for the sake of readability: '\( adb \)', for instance, is the sentence obtained by concatenating \( a \), the value of \( d \), and
\( b \).
we cannot add any more information while still having this guarantee. And so our local context algorithm predicts that the local context for '_-' in this sentence is $c \cap p$; generally, the local context for a left conjunct is the global context, plus the information expressed by the right conjunct. By exactly parallel reasoning, the local context for the right conjunct is the global context plus the information expressed by the left conjunct, since our algorithm is, again, perfectly symmetric.\(^{30,31}\)

I will treat local contexts as parameters of the index, accessible for semantic interpretation (in this I follow Klinedinst and Rothschild 2012). Thus, given the considerations just sketched, 'and' will have the semantics in (31):

\[(31) \quad \langle p \text{ and } q \rangle_{g,\kappa,w} = 1 \text{ if and only if } \langle p \rangle_{g,\kappa^p,w} = 1 \text{ and } \langle q \rangle_{g,\kappa^q,w} = 1\]

$\kappa^p$ is to be interpreted essentially as set intersection: $\kappa^p = \kappa \cap \{w : \langle p \rangle_{g,\kappa,w} = 1\}$.\(^{32}\)

### 1.5 Explaining Wittgenstein disjunctions

We're now in a position to see how the bounded theory makes sense of the embedding behavior of Wittgenstein sentences. Consider a Wittgenstein sentence with the form of (32) as evaluated in a context $c$:

\[(32) \quad \text{Might, p and not p}.\]

Consider any world $w$. Recall that the local context for a conjunct is the local context for the whole conjunction (in this case $c$) intersected with the content of the other conjunct (in this case, 'Not p'). Thus the local context for the left conjunct in (32) will be $\kappa^p$. The locality constraint in our semantics thus ensures that, in evaluating 'Might, p-', we do so relative to an accessibility relation with the property that only $\kappa^p$-worlds are in its image, and thus, in particular, that only $\kappa^p$-worlds are

\(^{30}\)I will leave the description of the results of Schlenker's framework at this informal level; proofs of all the results stated can be found in Schlenker (2009) or alternately as cited.

\(^{31}\)The local context for unembedded modals, on this theory of local contexts, is just the global context. Thus on the bounded theory, unembedded modals will be interpreted as quantifying over a subset of the global context set. I believe this is a broadly plausible constraint. One apparent counterexample comes from 'exocentric' readings of modals (Egan et al., 2005); see §1.7.3 below. Tense operators also shift local contexts, in somewhat surprising ways, as in: 'The keys might have been in the car, even though they turned out not to be' (see von Fintel and Gillies 2011; thanks to David Boylan). There is much work to be done to explore these facts, but I will set aside these complex issues for the present.

\(^{32}\)I will sometimes leave off the variable assignment subscript in what follows, for readability.
in its image. And so the left conjunct of (32) is guaranteed to be false, if the whole conjunction is well-defined: for any \( g \) relative to which the whole conjunction is well-defined, \( w \) will not be able to access a \( p \)-world under \( g(i) \), and so \( \neg \text{Might}_i p \) will be false, thanks to the core relational truth conditions in our semantics. And so the whole conjunction will be false.

This reasoning was perfectly general, and so it shows that any sentence with the form of (32) will be false at any world (provided it is well-defined). Importantly, since our local context algorithm is entirely symmetric, this account will extend immediately to Wittgenstein sentences which permute the order of the conjuncts, as in \( \neg p \) and \( \neg \text{Might}_i p \). And thus this reasoning shows that, on the bounded theory, any Wittgenstein sentence will be a classical contradiction in the sense that it is false at any index of evaluation where it is defined (I will use ‘contradiction’ in this sense throughout what follows).

Like most accounts, our account can thus explain the infelicity of Wittgenstein sentences. More significantly, however, this account straightforwardly extends to an explanation of the infelicity of Wittgenstein disjunctions. This is for the simple reason that, on our account of disjunction—as on any plausible account—the disjunction of two classical contradictions will always itself be a classical contradiction: a disjunction cannot be true at any world where both of its disjuncts are false.\(^{33}\) And so the bounded theory’s explanation of the infelicity of Wittgenstein sentences extends straightforwardly to an explanation of the infelicity of their disjunctions: Wittgenstein disjunctions, on the bounded theory, will be classical contradictions. (By contrast, of course, no parallel locality constraint applies to the interpretation of ‘knows’—thus explaining the divergence of Wittgenstein disjunctions from Moore disjunctions.)

Let us take a step back to appreciate how the bounded theory accomplishes this. Recall the

\[^{33}\text{To show this in more detail, note that in our framework for local contexts, the local context for a disjunct is the global context intersected with the complement of the other disjunct, and so disjunction gets the semantics in (33), where } \kappa^p_g = \kappa \setminus \{ w : [p]^{\kappa^g,w}_w = 1 \}; \]
\[^{33}\begin{align*}
[p \lor q]^{\kappa^g,w}_w &= 1 \iff [p]^{\kappa^g,w}_w = 1 \lor [q]^{\kappa^g,w}_w = 1 \\
[\text{I leave it to readers to confirm that, given this semantics, Wittgenstein disjunctions will be classical contradictions in my semantics. David Boylan points out to me an unintuitive feature of this semantics for disjunction, together with my semantics for epistemic modals, namely that } 'p \text{ or might}_i p' \text{ is predicted to be equivalent to } p. \text{ I'm unsure of how bad this consequence is: disjunctions like this are decidedly odd, and insofar as they are interpretable, I'm inclined to think that we interpret the 'or' as something like a particle signaling revision, rather than as an ordinary disjunction (on this use of 'or' see e.g. Szabolcsi 1997, 2015; Rawlins 2016).}
\end{align*}\]

36
puzzle that Wittgenstein disjunctions pose. On the one hand, "Might p" seems to be consistent with "Not p"; otherwise, "Might p" would entail p. On the other hand, the behavior of Wittgenstein disjunctions suggests that "Might p and not p" is not in fact consistent. My theory makes sense of these facts by predicting that "Might p" is indeed consistent with "Not p"—that is, both are true at the same worlds, relative to many variable assignments. But, if we evaluate their conjunction relative to a variable assignment which makes each conjunct true at a given world, the result will be undefined. The conjunction will be well-defined only if we evaluate it relative to an accessibility relation which renders the conjunction false at any world. And so, despite the joint consistency of the two conjuncts, the conjunction ends up being a contradiction.

The situation here is subtle, but the underlying mechanics of my solution are familiar: classical logic, plus the possibility of changes in interpretation arising from the structure of information in discourse. On my theory, again, "Might, p" is indeed consistent with "Not p". And the underlying logic is essentially classical, in the sense that classical inferences will preserve truth as long as they preserve definedness conditions (we can call this kind of classicality 'Strawson classicality', following von Fintel (1999)). But the caveat here is crucial. For any variable assignment g, there may be worlds where "Might, p" and "Not p" are both true, relative to g. But if this is so, then their conjunction—"Might, p and not p" or "Not p and might, p"—will be undefined, relative to g (and so conjunction introduction in this case will not preserve truth, since it will not preserve definedness). When we consider this conjunction, if we wish to maintain definedness, we will be forced to do so relative to a new variable assignment which makes the modal conjunct, and thus the whole conjunction, false at any world. In short: conjoining two sentences can change the way that we interpret those sentences. Interpreted separately, "Might, p" and "Not p" may well both be true relative to a given world and variable assignment; but their conjunction can never be interpreted in such a way that it is true, relative to any world and variable assignment.

34 More precisely, whenever Φ classically entails ψ relative to a variable assignment and context, then, in the bounded system, Φ entails ψ relative to that variable assignment and context as long as all the elements of Φ, and ψ, are well-defined relative to that variable assignment and context. All this also goes for K, the weakest normal modal system, which incorporates classical propositional logic. The proof of these facts turns on the observation that, apart from definedness conditions, my semantics is exactly like the semantics for K. Thus any sentence that is true in K will be true or undefined on my semantics, and any sentence false in K will be false in my semantics. And so, if a given inference rule preserves truth in K, it will preserve truth in my semantics as well, as long as all the premises and the conclusion are well-defined.
1.6 Order and the dynamics of information

At first glance, the bounded theory looks like a fairly close variant on the dynamic update approach to epistemic modals. In this section, I will explain why this is not so. The bounded theory departs in a fundamental way from the dynamic approach that it builds on, a departure which is essential to account for Wittgenstein disjunctions, and which cannot be satisfactorily replicated in the update (or domain) frameworks. In short—contrary to much recent work on the semantics of epistemic modals—I argue that the behavior of epistemic modals in Wittgenstein disjunctions does not tell against broadly relational approaches to their semantics, but in fact shows that we must adopt a broadly relational approach.

As we saw when introducing the data, the infelicity of Wittgenstein disjunctions does not depend at all on the order of the conjuncts in each disjunct: (8) and (10) are equally infelicitous.

(8) a. #Either I might win, but I won’t, or I might lose, but I won’t.
    b. #Might p and not p, or might q and not q.

(10) a. #Either I won’t win, but I might, or I won’t lose, but I might.
    b. #Not p and might p, or not q and might q.

This fact is crucial. The idea that the interpretation of epistemic modals under connectives depends on their local information is, again, not new. But, crucially, in essentially every extant version of this idea—most prominently in the update semantics—this dependency is asymmetric: the content of a left conjunct can influence the interpretation of a modal in a right conjunct, but not vice versa. This means that, as we saw above, views like this can predict the infelicity of sentences like (10), but not the infelicity of sentences like (8). By contrast, the bounded theory is, crucially, entirely symmetric: because the underlying theory of local contexts is symmetric, the content of a conjunct always limits the domain of quantification for a modal in the other conjunct, whatever their order in the sentence.

This symmetry is required to account for the full range of Wittgenstein disjunctions. These data, I think, thus resolve a longstanding empirical controversy about epistemic modals. Broadly dynamic
approaches claim that the interpretation of modals under connectives is sensitive to the intrasentential dynamics of information, but that that sensitivity is fundamentally asymmetric. Wittgenstein disjunctions show that there is something right in this approach and something wrong: the interpretation of epistemic modals is sensitive to the intrasentential dynamics of information, but not to the order of conjuncts. The intrasentential dynamics of information, at least when it comes to epistemic modals on either side of a conjunction, is a symmetric matter.

This does not mean that the dynamics of epistemic modality across discourses is itself symmetric; that, of course, is false. This suffices to account for observations like that of Groenendijk et al. (1996), who point to the differences between the following sequences:

(34) It might be raining outside. [...] It isn't raining outside.

(35) It isn't raining outside. [...] It might be raining outside.

There is certainly a difference to these sequences: (34) looks like an ordinary, 'monotonic' evolution of information; (35) looks like a change of mind. This shows that order matters across sequences of assertions. The bounded theory captures that fact straightforwardly: insofar as the first assertion in (35) changes the global context, the local context for the modal in (35) will differ from that for the modal in (34), since in calculating the local context for the modal in (34), the second sentence will not be taken into account (the symmetry in our local context algorithm is symmetry within sentences, not symmetry across discourses). But the fact that order across sequences of assertions matters does not show that there is semantically encoded asymmetry within sentences; and, I think, Wittgenstein disjunctions provide a powerful argument against that hypothesis.\footnote{See e.g. Yalcin 2012b, 2013 for similar conclusions.}

\footnote{I am thus committed to giving a broadly pragmatic explanation of the infelicity of a sequence like ‘It’s raining out. It might not be’, when the speaker gets no new evidence in between the sentences. A sequence like this doesn’t look so different from a Wittgenstein conjunction. While it might sometimes be interpreted as a conjunction, my contention is that, insofar as it is not interpreted that way, it is importantly different. Does this approach open me up to my own objection from Wittgenstein disjunctions? It does not, since, of course, we cannot embed a series of sentences in a disjunction. (There is such a thing as a particle ‘or’, which can act as a kind of speech act disjunction, as in “The keys are upstairs. Or, perhaps not” (see again citations in Footnote (33)). But I do not think we can use this to generate a problem here. A related point is that sometimes we can hear good readings of ‘It’s raining out, and it might not be’ (or the converse order), namely, when the speaker’s evidence changes across the course of the assertion. We can make sense of this in my framework by treating the conjunction here as something like a particle or speech act conjunction, rather than sentential conjunction, and parsing an assertion like this as having the form ‘It’s raining out. And, it might not be’. That this is the right approach is suggested, again, by the fact that conjunctions like this can’t be felicitously
Symmetry is thus an essential feature of the bounded theory. But it is not the only essential feature of the bounded theory. It is tempting to think that we could reproduce the effects of the bounded theory in a different framework, like the domain or update semantics, simply by adopting symmetric dynamic approaches within those frameworks. But this is mistaken: it turns out to be impossible to construct a satisfying symmetric dynamic domain or update approach. In the rest of this section, I will explain why this is. The upshot is that the relationality of the bounded theory is essential to that theory's strategy for making sense of Wittgenstein disjunctions.

Consider first what happens if we couple the domain semantics with dynamic and symmetric connectives along the following lines (the semantics for negation remains unchanged from above):\(^\text{37}\)

\[
\langle p \land q \rangle_{s,w} = 1 \text{ iff } [p]^{s,w} = 1 \text{ and } [q]^{s,w} = 1
\]

\[
\langle p \lor q \rangle_{s,w} = 1 \text{ iff } [p]^{s,w} = 1 \text{ or } [q]^{s,w} = 1
\]

Read \(s^p = s \cap \{ w : [p]^{s,w} = 1 \} \), and \(s^q = s \setminus \{ w : [p]^{s,w} = 1 \} \). If we adopt these connectives, then Wittgenstein disjunctions will indeed be predicted to be incoherent: no context will accept any Wittgenstein disjunction.

So far so good. But this approach runs into serious trouble elsewhere. Consider a disjunction like (38):

\[
\text{(38) } \begin{align*}
\text{(a) } & \text{The keys might be upstairs, or they might be downstairs.} \\
\text{(b) } & \text{Might } p \text{ or might } q.
\end{align*}
\]

Intuitively, a sentence like (38-a) should be accepted by an information state that includes worlds where the keys are upstairs and worlds where the keys are downstairs. But in the present framework, while ‘The keys might be upstairs, and they might be downstairs’ will be accepted in such an information state, (38-a) won’t be. To see this, consider a disjunction with the form of (38-b), and an information state \(s\) that includes both \(p\)-worlds and \(q\)-worlds. The local context for the first disjunct will be \(s \setminus \{ w : [\text{Might } q]^{s,w} = 1 \} \). But, since \(s\) includes a \(q\) world, ‘Might \(q\)’ will be\(^{37}\)

\(^{\text{37}}\)Cf. Klinedinst and Rothschild 2012, which couple the domain semantics with dynamic, but asymmetric, connectives.
true at every world in \( s \), and thus \( s \setminus \{ w : \text{Might } q \}^{\omega} = \{w \} \) is the empty set. So, in evaluating (38-b) at \( (s, w) \), we will evaluate the left disjunct at \( (\emptyset, w) \), and thus the left disjunct will be false, since the empty set does not include a \( p \)-world. Things are exactly parallel for the right disjunct. So the whole disjunction will be false at any world in \( s \), relative to \( s \). Thus \( s \) turns out not to accept (38-b). Worse, \( s \) does accept \( \neg \text{Not (might } p \text{ or might } q) \). Since \( s \) also accepts \( \text{Might } p \text{ and might } q \), it follows that \( \neg \text{Not (might } p \text{ or might } q) \) is predicted to be consistent with, and indeed jointly acceptable with, \( \text{Might } p \text{ and might } q \).

This is clearly unacceptable. It is a problem, moreover, that the bounded theory avoids: our symmetric entry for disjunction has none of these consequences. One option is to adopt a symmetric conjunction together with an asymmetric (or non-dynamic) disjunction. This would solve the present problem. But it is unsatisfying from an explanatory point of view. As a methodological principle, we do not want to stipulate the non-Boolean behavior of connectives on an ad hoc basis. Rather, we want to justify our choice in terms of an independently motivated algorithm, like the one the bounded theory adopts. I cannot see how we could justify, in a principled way, the coupling of a symmetric conjunction with an asymmetric disjunction.

Matters are even worse for an attempt to spell out a symmetric update semantics for epistemic modals. That semantics would go as follows (our semantics for negation, again, remains unchanged):

\[
(39) \quad c[p \text{ and } q] = c[p][q] \cap c[q][p]
\]

\[
(40) \quad c[p \text{ or } q] = c[\neg q][p] \cup c[\neg p][q]
\]

If we adopt these connectives, then we do indeed predict that Wittgenstein disjunctions will be incoherent: these sentences will invariably induce a ‘crash’ to the empty context.

---

38 The bounded theory ends up equivalent to this version of the domain theory if we say that the accessibility relation associated with modals has its image, at any world, identical to the local context, rather than a subset of its local context. So the present discussion brings out why it was crucial to formulate our locality constraint in terms of a subset, rather than identity, condition.

39 This objection is a variation on the ‘and*’ objection to Heim (1983)’s framework (see Soames 1989; Heim 1990): why should conjunction have the left-to-right asymmetry encoded in her framework, rather than a right-to-left asymmetry (or none at all)? See e.g. Stokke 2013 for a deflationary response.

40 This approach would be particularly unsatisfying insofar as data from presupposition projection show symmetries more clearly across disjunction than across conjunction.
But these connectives are empirically untenable. First, the symmetric disjunction faces the same problem raised for the symmetric domain approach, namely, that "Might p or might q" will not be accepted in a context which contains p-worlds and q-worlds, and in fact will always ‘crash’ such a context to the empty set. But here there is an even more basic problem with disjunction: (40) is no longer a truth-conditionally respectable semantics. It is not truth-conditionally equivalent to Boolean disjunction, but rather to exclusive disjunction. Suppose that s contains a world w where p and q are both true, and no other p-worlds or q-worlds. Intuitively, we want an update of s with [p or q] to result in \{w\}, since "p or q" is true at w. Instead, on this approach, it would result in ∅.

The issue is not just that we have picked the wrong symmetric dynamic connectives: as Rothschild (2015) argues, there is just no satisfactory symmetric dynamic semantics for disjunction. Again, we could adopt a symmetric conjunction and asymmetric disjunction. But this would be entirely unsatisfying. First, just as for the domain semantics, this is entirely unprincipled from an explanatory point of view. Second, a symmetric dynamic conjunction is untenable from the perspective of anaphora and presupposition projection—the two phenomena that dynamic semantics was designed to make sense of. The symmetric dynamic conjunction in (39) would wrongly predict that anaphora across conjunction is order-invariant; but this is clearly false. And the symmetric dynamic conjunction would predict that presuppositions always project out of conjuncts. While there is some controversy about whether presupposition projection is symmetric or asymmetric, as I discuss in a moment, no one thinks that presuppositions always project out of conjunctions. A symmetric dynamic semantics is thus even worse off than a symmetric approach to the domain semantics.

This is not an argument against CCP-style dynamic semantics in general. Rather, this is an argument that we cannot exploit the dynamics which dynamic semantics use to make sense of anaphora and presupposition in order to also make sense of epistemic modals, as the update semantics aims to do. If we want a symmetric theory of epistemic modals, then we can only build one in the relational

---

41We’ll first update the context with "Not might q", getting the empty set (since there is a q-world in the context), then update with "Might p", which of course leaves us with the empty set; simultaneously we’ll update the context with "Not might p", again getting the empty set (since there is a p-world in the context), then update it with "Might q", leaving us with the empty set; taking their union, we again have the empty set.

42Thanks to Daniel Rothschild, who pointed this out to me.
framework, and not in the domain or update framework. And thus, since I believe that the embedding behavior of epistemic modals in Wittgenstein disjunctions show that we need a symmetric approach to epistemic modals, this also shows that we need to take a broadly relational approach to epistemic modals.

Let me close by commenting on where all of this leaves us with regard to the dynamics of information in natural language. The idea that local contexts must be calculated symmetrically, and only symmetrically, runs counter to nearly all work on the role of local contexts in presupposition projection, which assumes that there are genuine asymmetries in presupposition projection. Symmetric algorithms for calculating local contexts have indeed been proposed, for instance in Schlenker 2008a, 2009; Rothschild 2015). But Schlenker proposes that the symmetric algorithm is recruited only when the asymmetric algorithm seems to be going wrong. Rothschild suggests that a left-to-right parse and a right-to-left parse are more or less on a par, and that both are typically available. Neither of these approaches would suffice to account for Wittgenstein disjunctions. If both algorithms were available, we would expect that interlocutors would simply use an algorithm which would render Wittgenstein sentences felicitous (and there will always be such an algorithm, on either of these theories).

If my argument from Wittgenstein disjunctions for a symmetric approach to local contexts is convincing, we are left with two options. The first is to conclude that epistemic modals and presupposition projection depend on different systems of local contexts. This is certainly an option that we should take seriously. But it is also unattractive for reasons of theoretical parsimony: it would be more elegant to posit that presupposition and epistemic modality both operate on the same underlying system of local information. If we take this route, then we will have to attribute observed asymmetries in projection to the interference of independent factors having to do with the way that we process sentences across time. There is much work to be done to see if an approach like this is plausible.43 One important point to note if we go this way is that we could not then rely on local

43For an idea long those lines, see Rothschild 2008a. For some recent experimental evidence both for and against an approach like this, see Chemla and Schlenker 2012; Schwarz 2015; Mandelkern et al. 2017. Straightforwardly applying Schlenker’s symmetric algorithm to presupposition projection faces a serious problem, pointed out by Rothschild (2008b); Beaver (2008), that the same presupposition trigger appearing on both sides of a connective is predicted to be able to “cancel itself out”. See Schlenker (2008b) for a possible, if not altogether satisfying, response.
contexts for explaining systems which are undeniably asymmetric, like anaphora.

1.7 Empirical reach

The bounded theory accounts for Wittgenstein disjunctions, something that almost no other theory of epistemic modals can do. But Wittgenstein disjunctions are not the only game in town. I will conclude my exposition of the bounded theory by showing that it accounts not just for Wittgenstein disjunctions, but also for the embedding behavior of epistemic modals in a wide variety of further environments. Moreover, the bounded theory does so in a principled and predictive way—in virtue of a simple and compact generalization about the derivation of local contexts, and their influence on epistemic modals.

The discussion in this section will remain fairly informal. In Appendix A, I compile all of the semantic entries relevant to evaluating the claims I make here; interested readers can use those entries to verify my claims.

1.7.1 Attitudes and conditionals

Yalcin (2007) points out the infelicity of Wittgenstein sentences under ‘Suppose’ or ‘If’:

(41)  a. #Suppose it’s raining and it might not be.
     b. #Suppose p and might not p.

(42)  a. #If it’s raining and it might not be, then...
     b. #If p and might not p, then...

This is surprising from the point of view of the standard (non-local) relational theory, on which a sentence like (41-a) should mean something like ‘Suppose it’s raining and, for all we know, it isn’t’, which is perfectly felicitous. But our local relational semantics, plus our theory of local contexts, provides a simple and unified account of the infelicity of these sentences. Since Wittgenstein sentences, on our theory, are contradictions, they will embed just like ordinary contradictions. So

\[\text{Yalcin focuses on variants with the modal in the right conjunct, which he calls 'epistemic contradictions', but his point extends to order permutations.}\]
a sentence like (41-a) will be equivalent to a command to make your suppositions inconsistent (I spell this out more carefully in a footnote), and thus will be infelicitous for the same reason that

Suppose it’s raining and it isn’t raining is infelicitous: because there is something wrong with commanding someone to make their suppositions inconsistent. Likewise, (42-a) will be infelicitous for the same reason that If it’s raining and it isn’t raining, then... is infelicitous: because, again, it amounts to the supposition of an impossibility. And since everything in the bounded theory is perfectly symmetric, these explanations will extend immediately to account for variants of these sentences which change the order of the conjuncts (something which does not affect intuitions about their meaning), as well as to variants on (41-a) with a high-scope conjunction, as in (45);\(^4\) Consider the indicative variant of (41-a), -A supposes (p and might not p). In our system of local contexts, assuming a standard Hintikka semantics for ‘suppose’ as our underlying “classical” semantics, and where \(S_{A,w}\) is the set of worlds compatible with A’s suppositions in \(w\)—the local context introduced by ‘supposes’ in context c is \(\bigcup_{w \in c} S_{A,w}\) i.e. the set of worlds compatible with what A supposes, for all that is accepted in c. (Schlenker 2009) actually suggests that the local context is a world-relative matter, and that at \(w\), ‘A supposes’ introduces the local context \(S_{A,w}\). I believe that to get this result, one has to modify his algorithm in ways that are problematic elsewhere, though for our purposes, however, either will suffice. Thanks to Daniel Rothschild for discussion.) For readability, I’ll abbreviate \(\bigcup_{w \in c} S_{A,w}\) as \(S\).

So our semantics for ‘suppose’ will run: \([A \text{ supposes } q]^{g,c,w} = 1 \text{ if } \forall w' \in S_{A,w} : [q]^{g,S,w'} = 1\). Now consider any world \(w\). Given our semantics for ‘suppose’, ”A supposes (p and might not p)” is true at \(w\) and the relevant variable assignment, \(g\), just in case \(\forall w' \in S_{A,w} : [p \text{ and } \text{ might } \text{ not } p]^{g,S,w'} = 1\). Given our semantics for conjunction, and using obvious abbreviations, this holds just in case \(\forall w' \in S_{A,w} : [p]^{g,S,w'} = 1 \wedge [\text{ might } \text{ not } p]^{g,S,w'} = 1\). Suppose these conditions are met and \(S_{A,w}\) is non-empty. Then for some \(w' \in S_{A,w}\), \(\text{ might } \text{ not } p\) \(= 1\), and so \(w'\) can access a \(p\)-world under \(g(i)\). But we know that \(g(i)\) will only be able to access \(S^p\)-worlds, and thus \(p\)-worlds, under \(g(i)\), thanks to the locality constraint, and so \([\text{ might } \text{ not } p]^{g,S,w'} = 0\). Thus we can conclude that \(S_{A,w}\) is empty, after all. Since \(w\) was chosen arbitrarily, it follows that this claim is equivalent to the claim that A’s suppositions are inconsistent: i.e. to the claim ”A supposes \(p\) and \(\text{ not } p\)”.

There is more to say about what is wrong with supposing an impossibility—many have thought the explanation is that conditionals presuppose their antecedents to be compatible with the global context, a presupposition which of course could not be met in this case (see e.g. Stalnaker 1975; von Fintel 1998b; Gillies 2009), and some have proposed that there is a corresponding presupposition that attitude verbs quantify over a non-empty set of worlds (see Ninan 2016)—but whatever we say there, this fact will explain the infelicity of the former.

Dorr and Hawthorne (2013) claim otherwise, but I believe the cases they give can be explained in terms of a change in time index between the modal and non-modal conjuncts. I am ignoring tense here, but implicitly assuming that the modal and non-modal conjuncts are evaluated relative to the same time.

The crucial point in showing this is that the local context for the right conjunct of (45-b), according to our algorithm, will be \(S^p\).

An interesting further prediction of the bounded theory is the following. Compare (43) with (44) (parentheses disambiguating scope):

(43) a. ??I believe it’s raining, but I believe it might not be raining.
   b. ??(I believe p), and (I believe might not p).

(44) a. I believe it’s raining, but it might not be raining.
   b. (I believe p), and (might not p).

There is a marked difference in felicity between (43)—which is felt to ascribe contradictory beliefs to the speaker—versus (44), which is not. This is predicted in the present system, on which (43) will indeed ascribe contradictory beliefs to the speaker, for the same reason that the variant with ‘suppose’ does; while (44) will not, since the modal is not in the

\(^{4a}\) Consider the indicative variant of (41-a), ”A supposes (p and might, not p). In our system of local contexts, assuming a standard Hintikka semantics for ‘suppose’ as our underlying “classical” semantics, and where \(S_{A,w}\) is the set of worlds compatible with A’s suppositions in \(w\)—the local context introduced by ‘supposes’ in context c is \(\bigcup_{w \in c} S_{A,w}\) i.e. the set of worlds compatible with what A supposes, for all that is accepted in c. (Schlenker 2009) actually suggests that the local context is a world-relative matter, and that at \(w\), ‘A supposes’ introduces the local context \(S_{A,w}\). I believe that to get this result, one has to modify his algorithm in ways that are problematic elsewhere, though for our purposes, however, either will suffice. Thanks to Daniel Rothschild for discussion.) For readability, I’ll abbreviate \(\bigcup_{w \in c} S_{A,w}\) as \(S\).

So our semantics for ‘suppose’ will run: \([A \text{ supposes } q]^{g,c,w} = 1 \text{ if } \forall w' \in S_{A,w} : [q]^{g,S,w'} = 1\). Now consider any world \(w\). Given our semantics for ‘suppose’, ”A supposes (p and might, not p)” is true at \(w\) and the relevant variable assignment, \(g\), just in case \(\forall w' \in S_{A,w} : [p \text{ and } \text{ might, not } p]^{g,S,w'} = 1\). Given our semantics for conjunction, and using obvious abbreviations, this holds just in case \(\forall w' \in S_{A,w} : [p]^{g,S,w'} = 1 \wedge [\text{ might, not } p]^{g,S,w'} = 1\). Suppose these conditions are met and \(S_{A,w}\) is non-empty. Then for some \(w' \in S_{A,w}\), \(\text{ might, not } p\) \(= 1\), and so \(w'\) can access a \(p\)-world under \(g(i)\). But we know that \(g(i)\) will only be able to access \(S^p\)-worlds, and thus \(p\)-worlds, under \(g(i)\), thanks to the locality constraint, and so \([\text{ might, not } p]^{g,S,w'} = 0\). Thus we can conclude that \(S_{A,w}\) is empty, after all. Since \(w\) was chosen arbitrarily, it follows that this claim is equivalent to the claim that A’s suppositions are inconsistent: i.e. to the claim ”A supposes \(p\) and \(\text{ not } p\)”.

\(^{4b}\) There is more to say about what is wrong with supposing an impossibility—many have thought the explanation is that conditionals presuppose their antecedents to be compatible with the global context, a presupposition which of course could not be met in this case (see e.g. Stalnaker 1975; von Fintel 1998b; Gillies 2009), and some have proposed that there is a corresponding presupposition that attitude verbs quantify over a non-empty set of worlds (see Ninan 2016)—but whatever we say there, this fact will explain the infelicity of the former.

\(^{4c}\) Dorr and Hawthorne (2013) claim otherwise, but I believe the cases they give can be explained in terms of a change in time index between the modal and non-modal conjuncts. I am ignoring tense here, but implicitly assuming that the modal and non-modal conjuncts are evaluated relative to the same time.

\(^{4d}\) The crucial point in showing this is that the local context for the right conjunct of (45-b), according to our algorithm, will be \(S^p\).

\(^{4e}\) An interesting further prediction of the bounded theory is the following. Compare (43) with (44) (parentheses disambiguating scope):

(43) a. ??I believe it’s raining, but I believe it might not be raining.
   b. ??(I believe p), and (I believe might not p).

(44) a. I believe it’s raining, but it might not be raining.
   b. (I believe p), and (might not p).

There is a marked difference in felicity between (43)—which is felt to ascribe contradictory beliefs to the speaker—versus (44), which is not. This is predicted in the present system, on which (43) will indeed ascribe contradictory beliefs to the speaker, for the same reason that the variant with ‘suppose’ does; while (44) will not, since the modal is not in the
(45)  a. Suppose it's raining and suppose it might not be raining!
    b. Suppose p and suppose might not p!

1.7.2 **Quantifiers**

One of the selling points of Schlenker’s algorithm for calculating local contexts—and one of the
main ways in which it improves on earlier informal pragmatic work—is that it makes precise predic-
tions about the local contexts for material embedded under quantifiers. Together with our semantics
for modals, this approach to local contexts makes sense of much of the puzzling and subtle behavior
of epistemic modals under quantifiers.\(^\text{50}\)

Consider first modals under ‘some’. Sentences with the form of (46) and (47) are infelicitous,
as Groenendijk et al. (1996) and Aloni (2001), respectively, observed:\(^\text{51}\)

(46)  a. #Someone hiding in the closet might not be hiding in the closet.  [Groenendijk et al.
        1996]
    b. #Some(p)(might, not p).

(47)  a. #Someone who might not be hiding in the closet is hiding in the closet.
    b. #Some(might, not p)(p).

Assuming, as is standard, that \(\exists Some(p)(q)\) just says that some individual is both \(p\) and \(q\), this is
surprising from the point of view of the standard (non-local) relational semantics, on which (46-b)
or (47-b) will be true just in case there is some individual \(a\) such that \(a\) is \(p\), and it is compatible
with what we (the relevant agents) know that \(a\) is not \(p\). It should be very easy for these conditions
to obtain, if, say, we know of two people that one of them is \(p\) and one of them isn’t, but we don’t
know which is which.

Assuming this standard meaning for ‘some’, our local context algorithm predicts that the local

---

\(^\text{50}\)For a first intimation of the puzzles arising from epistemic modals under quantifiers, see Beaver 1994; for further dis-
cussion, see Groenendijk et al. 1996; Gerbrandy 1998; Aloni 2000, 2001; Yalcin 2015; Rothschild and Klinedinst 2015;
Ninan 2017. Note that roman letters in this discussion stand for open sentences, and italic letters for the corresponding
predicates, obtained by abstracting over unbound variables.

\(^\text{51}\)Yalcin (2015) reports that the latter order variant is somewhat preferred to the former; more on this below.
context for \( p \) in \( \forall \)Some(p)(q)\( \land \) entails \( q \).\(^{52}\) To see this, note that a sentence of this form will always be equivalent to \( \forall \)Some(p)(p and q)\( \land \) (e.g., ‘Some male is tall’ is equivalent to ‘Some male is tall and male’). Likewise, the local context for \( q \) will entail \( p \): \( \forall \)Some(p)(q)\( \land \) and \( \forall \)Some(p and q)(q)\( \land \) will always be equivalent (e.g., ‘Some male is tall’ and ‘Some tall male is tall’ are equivalent). This means, very roughly, that \( \forall \)Some(p)(\text{might}_i \not p)\( \land \) will say of some individual \( a \) that it is \( p \), and that there is some accessible world where \( a \) is \( \bar{p} \). But, since the local context for the nuclear scope (the second argument of ‘some’) entails \( p \), worlds where \( a \) is \( p \) will only be able to access other worlds where \( a \) is \( p \) under the relevant accessibility relation \( g(i) \), and so \( \forall \)\text{Might}_i \not p\( \land \) will be false of \( a \) at any world where \( p \) is true of \( a \). The upshot is that sentences with the form of (46-b) or (47-b) will be contradictions (I make this more precise in a footnote).\(^{53}\)

A similar point concerns definite descriptions. As Aloni (2001) observed, sentences like (49) are infelicitous:

\[
\text{(49) a. } \#\text{The biggest flea might be the smallest flea.} \quad [\text{Aloni 2001}]
\]
\[
\text{b. } \#\text{The(p)(might}_i \not p).\]

This is, again, puzzling on the standard (non-local) relational semantics, on which (49-a) should mean something like: there is a unique biggest flea \( a \), and, for all we know, \( a \) is the smallest flea. The bounded theory explains these cases in essentially the same way it explains the infelicity of the corresponding sentence with ‘some’ in (46-b). The local context for \( q \) in \( \forall \)The(p)(q)\( \land \) entails \( p \),

\(^{52}\)The notion of entailment is generalized here to types other than propositions; see again Schlenker 2009 for formal details. When I say in this section that the domain of quantification is limited to \( q \)-worlds, this is slightly periphrastic: as the semantics in (48) makes clear, the relevant local context will in fact vary with the free variable.

\(^{53}\)Given our local context calculations, our semantics for ‘some’ will be the following (where \( p \) and \( q \) have the type of functions from indices to functions from individuals to truth values):

\[
[\text{Some}(p)(q)]^{\varphi_0,\varphi_1,\varphi_2} = 1 \iff \exists x : [p]^{\varphi_0,\varphi_1,\varphi_2}(x'w') = 1, [q]^{\varphi_0,\varphi_1,\varphi_2}(x'w') = 1, \varphi_1(x') = 1
\]

Now substitute \( \forall \text{Might}_i \not p \) for \( q \) in (48), so that we have a sentence with the form of (46-b). Given the standard Tarski semantics for \( \exists \), the right-hand side of this biconditional will then be true iff for some individual \( a \), \( [p]^{\varphi_0,\varphi_1,\varphi_2}(w'x') = 1 \), \( [q]^{\varphi_0,\varphi_1,\varphi_2}(w'x') = 1 \), \( w'x' \) holds of \( a \), and \( [\text{Might}_i \not p]^{\varphi_0,\varphi_1,\varphi_2}(w'x') = 1 \), \( w'x' \) holds of \( a \). Assume these two conditions hold for some \( a \) at some \( w \). Then, thanks to locality, under the accessibility relation \( g(i) \), \( w \) can only access worlds in the local context, and thus \( w \) can only access worlds where \( a \) is \( p \). It follows from the core relational part of our semantics that the right conjunct is false, since \( w \) cannot access any world where \( p \) is false of \( a \). So these two conditions cannot, after all, both hold for any \( a \) at \( w \). This shows that there is no witness to the truth of (46-b) at \( w \), and thus that (46-b) is false at \( w \). And since \( w \) was chosen arbitrarily, this shows that (46-b) is false at any world where defined: that is, (46-b) is a contradiction. Our truth conditions in (48) are entirely symmetric, so exactly the same reasoning will show that (47-b) is, likewise, a contradiction.
according to our algorithm; we can see this by noting the felt equivalence between \( \forall \text{The}(p)(q) \) and \( \forall \text{The}(p)(p \land q) \) (for instance ‘The dog is brown’ is felt to be equivalent to ‘The dog is brown and a dog’).\(^\text{54}\) Then a sentence with the form of (49-b) will be felt to say, roughly: there is a unique \( p \)-individual \( a \), and there is some accessible world (under \( g(i) \)) where \( a \) is \( \bar{p} \). But because the local context for the nuclear scope entails \( p \), the accessibility relation \( g(i) \) will only be able to access worlds where \( a \) is \( p \). And so no world will be able to access a world under \( g(i) \) where \( a \) is \( \bar{p} \). (49-b) will thus again be a contradiction. (Readers can confirm this on the basis of the semantics for ‘the’ given in Appendix A.)

Let me point to a final, intriguing set of predictions of the bounded theory. Yalcin (2015) (crediting Declan Smithies) points out that sentences like (50) are felicitous.

\[
\begin{align*}
(50) & \quad \text{a. Not everyone who might be sick is sick.} \\
& \quad \text{b. Not every}(\text{might}, p)(p).
\end{align*}
\]

This is surprising, since, on the face of it, (50-b) looks logically equivalent to sentences with the form of (47-b), which, as we saw above, are infelicitous:

\[
\begin{align*}
(51) & \quad \text{a. } \#\text{Someone who might be sick is not sick.} \\
& \quad \text{b. } \#\text{Some}(\text{might, } p)(\text{not } p).
\end{align*}
\]

It looks like we can move between (50) and (51) simply by permuting negation. So why is (50) substantially more felicitous than (51)?

It is not immediately clear how our theory helps here. If we treat ‘every’ as the familiar universal quantifier from first order logic, then the local context for \( p \) in \( \forall \text{Every}(p)(q) \) will entail \( \bar{q} \), since \( \forall x : p(x) \rightarrow q(x) \) is logically equivalent to \( \forall x : (p(x) \land \neg q(x)) \rightarrow q(x) \). If we go this way, then \( \forall \text{Every}(\text{might}, p)(p) \) will be a tautology: since the modal in its restrictor (its first argument) will only be able to access \( \bar{p} \)-worlds under \( g(i) \), the restrictor will be true of no individuals at any context world, and thus the sentence will be trivially true at any context world. And so its negation in (50-b) is (wrongly) predicted to be a contradiction, like (51).

\(^\text{54}\)This result will follow from our algorithm under any reasonable assumptions about the semantics of ‘the’.
But if we take a more sophisticated approach to the semantics of 'every', then our theory does not predict that the local context for \( p \) in \( \forall \text{Every}(p)(q) \) entails \( \bar{q} \). To see the motivation for this, note that, contrary to our predictions if we treat 'every' as \( \forall \), \( \forall \text{Every}(p)(q) \) is intuitively not equivalent to \( \forall \text{Every}(p \text{ and not } q)(q) \), as witnessed, for instance, by (52) and (53):

\[(52) \quad \text{Every dog is in the genus } canis.\]

\[(53) \quad \#\text{Every dog which isn’t in the genus } canis \text{ is in the genus } canis.\]

Again, these are predicted equivalent if 'every' is treated as \( \forall \). We can break this equivalence, however, if we follow much work in linguistic theory in positing that the meaning of 'every' diverges from that of \( \forall \) in that the former, but not the latter, presupposes that the extension of its restrictor is non-empty.\(^{55}\) This presupposition accounts for the felt difference between (52) and (53): the former presupposes that there is some dog, and asserts (truly) that anything which is a dog is in the genus \( canis \); the latter presupposes (falsely) that there is a dog not in the genus \( canis \), while simultaneously asserting that anything which is a dog is in the genus \( canis \). The total content of each thus ends up being quite different: the former can be (in fact, is) true, whereas the combined presupposed and asserted content of the latter cannot be true.

If we take into account the existence presupposition of 'every' in calculating the local context for its restrictor, then \( \forall \text{Every}(p)(q) \) and \( \forall \text{Every}(p \text{ and not } q)(q) \) are not invariably equivalent, and thus our algorithm predicts that the local context for \( p \) will not entail \( \bar{q} \), and turns out just to be the global context.\(^{56}\) Thus, in particular, in \( \forall \text{Every} (\text{might}_i p)(p) \), as assessed at context \( c \), the modal's local context will be \( c \). And so we will interpret the embedded modal just as we would interpret an unembedded modal. A sentence with the form \( \forall \text{Every} (\text{might}_i p)(p) \) will thus have the meaning that we would naively expect it to have: it will simply say that anyone who is \( p \) in some accessible world, is in fact \( p \). This sentence will be neither a contradiction nor a tautology, and so neither will its negation: \( \forall \text{Not every} (\text{might}_i p)(p) \) will mean that someone who is \( p \) in an accessible world is in fact \( \bar{p} \). This is a perfectly coherent meaning.

\(^{55}\)See Heim and Kratzer (1998, §§6.8) for discussion and references.

\(^{56}\)The proof is essentially identical to a parallel result for presuppositional conditionals in Mandelkern and Romoli 2016.
If the local context for the restrictor of ‘every’ is blind to the information carried by its nuclear scope, then it may be objected that we will wrongly predict that a sentence with the form of \( \forall \text{Everyone}(\text{might}, p)(\neg p) \) is consistent. This appears to be wrong, as witnessed by the infelicity of e.g. (54):

(54)  #Everyone who might be sick is not sick.

But we can easily proffer an alternate explanation of the infelicity of sentences like this. ‘Every’, again, is associated with a non-emptiness presupposition, which predicts that (54) presupposes that someone might be sick. But (54) simultaneously asserts that no one who might be sick is sick. We cannot consistently accept both the content of its presupposition and its asserted content. 57 (A similar point can be made regarding sentences with the form \( \forall \text{The}(\text{might}, \neg p)(p) \).) That this is the correct explanation of the infelicity of (54) is brought out by its contrast with (57-a), which seems like it should mean roughly the same thing as (54), but is felicitous. 58

(57)  a. No one who might be sick is sick.

b. No one (might, p)(p)

The crucial difference between (57) and (54) is that ‘no one’ lacks the existential presupposition of ‘everyone’—meaning that the asserted content of (57) is predicted to be consistent with its presupposed content, and thus the overall package will be perfectly consistent.

Let me generalize the reasoning here. It is generally accepted that all (or nearly all) natural language quantifiers are conservative, in the sense that any quantificational structure \( \forall Q(a)(b) \) is equivalent to \( \forall Q(a)(a \text{ and } b) \). 59 This means that, for any quantificational structure \( \forall Q(a)(b) \), on

57 There is some subtlety about how to spell this out, which depends on the theory of presupposition; however we go, there should be a satisfying way of spelling this out.
56 Irene Heim points out to me that this approach predicts that the infelicity of (54) will not always persist in embedded environments. This approach seems to be borne out; thus (55) strikes me as much more felicitous than (54) (changing ‘is not sick’ to ‘is well’ for readability):

(55)  Not everyone who might be sick is well.

(55) strikes me as more or less equivalent to (56), which is perfectly felicitous:

(56)  Someone who might be sick is sick.

Judgments here are, however, somewhat subtle.
59 See Barwise and Cooper 1981.
our algorithm, the local context for b will entail a. But the converse does not invariably hold: the local context for a will only entail b when the quantifier is “conservative in its restrictor”, that is, only when \( \forall Q(a)(b)^\forall \) is invariably equivalent to \( \forall Q(a \text{ and } b)(b)^\forall \). That equivalence sometimes holds, as in the case of ‘some’. But it often fails to hold, for various reasons—for instance, due to a presupposition of ‘Q’ (as we’ve just seen for ‘every’), or due to the lexical semantics of ‘Q’ (as for ‘most’; see §1.8.2). When that equivalence is blocked, we expect to find asymmetries like those observed between (47) and (50).

This is a striking prediction; it is borne out in the data we’ve seen so far, but needs to be more thoroughly explored. I cannot undertake a systematic exploration here, but let me note one central upshot of this account. Many take ‘weak’ (i.e. generally non-presuppositional) quantifiers like ‘some’ to have presuppositional uses. Depending how we analyze this—and how this interacts with our local context algorithm—this may then lead us to predict that there are uses of weak quantifiers where they pattern like strong (presuppositional) quantifiers with respect to embedded modals. This, in turn, may explain Yalcin (2015)’s observation that some speakers find \( \forall \text{Some}(\text{might}_t p)(\text{not } p)^\forall \) to be more felicitous than \( \forall \text{Some}(\text{not } p)(\text{might}_t p)^\forall \). If a quantifier like ‘some’ has presuppositional uses, then the bounded theory predicts that, on its presuppositional uses, \( \forall \text{Some}(\text{might}_t p)(\text{not } p)^\forall \) will be predicted to be consistent. By contrast, \( \forall \text{Some}(\text{not } p)(\text{might}_t p)^\forall \) will never have a consistent use. This furnishes a possible explanation for the subtle contrast between these two.

There is much more to explore here, but this discussion should suffice to give a sense of the empirical reach of the bounded theory.  

\footnote{Thanks to Kai von Fintel for discussion; see von Fintel 1998a.}

\footnote{Let me point out two areas which need further exploration. Daniel Rothschild points out to me that quantifiers like ‘a few’ or ‘at least three’ look to be conservative in their restrictor; but nonetheless, a sentence like ‘A few people who might be sick aren’t sick’ sounds acceptable. This is indeed puzzling for my theory. The question is whether, again, insofar as these are felicitous, it is because we are bringing out a presuppositional reading of the quantifier. Second, Ninan (2017) points out the oddness of conjunctions like ‘Anyone might be the winner, but of course the loser can’t be the winner’. My theory—like one that Ninan is arguing against—predicts sequences like this to be consistent. The challenge for my theory is to see if a pragmatic explanation for the \textit{prima facie} oddness of a sentence like this can be proffered; I am hopeful that one can.}
1.7.3 Non-default readings

Let me close this section by briefly noting that many of the sentences which I have marked with a ‘#’ can be used in ways that are not obviously terrible. Suppose ornithology students are learning to identify birds and the instructor says (58):

(58) Whenever a bird is shown that isn’t a sparrow, but might be, the students hesitate. (Kratzer and Phillips, 2017)

Given a plausible semantics for ‘whenever’, (58) will be predicted, on the bounded theory, to be equivalent to (59):

(59) #Whenever a bird is shown that isn’t a sparrow and is a sparrow, the students hesitate.

But obviously (58) manages to communicate something different than (59); and while (58) is by no means impeccable, it certainly does not sound as strange as (59).

Similar readings can be brought out in other cases; the generalization seems to be that the hash-marks used throughout the paper identify a strong default reading, but that, provided sufficient contextual set-up, improved readings are sometimes available (cf. Egan et al. (2005)’s ‘exocentric’ readings). There are at least two ways to make sense of this within the present framework. The first is to follow the strategy of Stephenson 2007b; MacFarlane 2011 and to maintain that, on the non-default readings, we interpret the epistemic modals in question as embedded under a covert operator which shifts the local context—so that, for instance, (58), on its acceptable reading, is parsed as ‘Whenever a bird is shown that isn’t a sparrow, but, for all the students know, might be, the students hesitate’. ‘For all the students know’ will shift the local context relative to which ‘might’ is embedded to the worlds compatible with the students’ knowledge, explaining why (58) communicates what it does, and is not equivalent to (59). The second option is to hold that the calculation of local contexts which I have assumed is itself only a strong default which can be violated given sufficient contextual set-up. Both these options seem like viable routes for making sense of non-default readings of the sentences under discussion. Much more work needs to be done.

---

62There is a parallel here to local accommodation in the theory of presupposition projection; see Heim 1983.
to spell each out and decide between them; I will not try to do so here.

1.8 Alternative solutions

This completes my exposition of the bounded theory. In closing, I will compare the bounded theory to the only other extant approaches I know of which can account for the infelicity of Wittgenstein disjunctions, pointing to drawbacks of each which the bounded theory avoids.63

1.8.1 Enriched update semantics (i): Yalcin 2015

Various enrichments to the update semantics have been proposed in recent years. Two account for Wittgenstein disjunctions. The first, given in Yalcin (2015), augments a standard (asymmetric) update semantics (relevantly the same as the one given in §2.2.2 above) with the following constraint

\[ \text{(60)} \quad [\text{Might, } p]^{g, \kappa, w} \]

a. defined only if \( \forall w' \in \kappa : g(i)(w') \subseteq \kappa; \)

b. if defined, true iff \( \exists w' \in g(i)(w) : [p]^{g, \kappa, w} = 1. \)

This approach is very similar to the one in the text; it predicts that Wittgenstein sentences are contextual contradictions, rather than contradictions simpliciter. It thus provides an adequate account of the data discussed here, and is prima facie to be preferred since it is slightly weaker. I have gone with the stronger constraint in the main text in large part because it is simpler to present. The choice between the two is subtle. On the one hand, I believe the stronger constraint makes better predictions about the behavior of epistemic modals in the consequents of conditionals (predictions essentially equivalent to those e.g. of Kratzer 1991). On the other hand, the stronger constraint makes slightly strange predictions about disjoined 'might's. Consider a context which contains \( p- \) and \( p \)-worlds, with \( g(i) \) an accessibility relation which is universal within the context. Then the stronger constraint predicts "Might, \( p \) or might, not \( p \)" is false throughout the context. This is prima facie wrong. On the other hand, we can get a true reading of "Might \( p \) or might, not \( p \)" under these assumptions by co-indexing the 'might's differently, so that we have e.g. "Might, \( p \) or might, not \( p \)", and then making \( g(j) \) the empty accessibility relation. This is a bit ad hoc, but shows that this consideration does not tell conclusively against the stronger constraint; in light of the behavior of modals under conditionals, I'm inclined towards the stronger constraint, but deciding between them is a task for another day.

A different variant replaces universal quantification with bounded existential quantification, so that we replace (60-a) with: 'defined only if \( \kappa \neq \emptyset \rightarrow (\exists w \in \kappa : g(i)(w) \subseteq \kappa)' . On the one hand, this constraint predicts that Wittgenstein disjunctions are contextually consistent. However, it also predicts that no non-empty context is a fixed point for a Wittgenstein disjunction, i.e. there is no Wittgenstein disjunction \( d \) and non-empty context \( c \) such that \( c \cap \{ w : [d]^{g, c, w} = 1 \} = c \). It is not clear that this fact explains the infelicity of Wittgenstein disjunctions. One way around this is to appeal to a kind of prospective interpretation (see Stalnaker 2014 and Chapter 2 of this dissertation). I am sympathetic to this line, but do not want to take on this commitment for present purposes. A different, more economical variant of my approach has embedding operators shift just one parameter, a variable assignment, by shifting all variables with the type of an accessibility relation an accessibility relation which satisfies a locality constraint. This would bring our proposal closer to the mechanics of the proposal for the semantics of conditionals given in von Fintel 1994, but it would predict that accessibility relations for other flavors of modals (deontic, circumstantial) are shifted like those for epistemic modals, which, as I discuss in the conclusion, does not seem correct.

63In addition to exploring these variants, it is worth pointing to a few mild variants of the bounded theory. Most interestingly, a slight weakening of my locality constraint replaces universal quantification with universal quantification bounded to the local context worlds, so that we have:

\[ \text{(60)} \quad [\text{Might, } p]^{g, \kappa, w} \]

a. defined only if \( \forall w' \in \kappa : g(i)(w') \subseteq \kappa; \)

b. if defined, true iff \( \exists w' \in g(i)(w) : [p]^{g, \kappa, w} = 1. \)

This approach is very similar to the one in the text; it predicts that Wittgenstein sentences are contextual contradictions, rather than contradictions simpliciter. It thus provides an adequate account of the data discussed here, and is prima facie to be preferred since it is slightly weaker. I have gone with the stronger constraint in the main text in large part because it is simpler to present. The choice between the two is subtle. On the one hand, I believe the stronger constraint makes better predictions about the behavior of epistemic modals in the consequents of conditionals (predictions essentially equivalent to those e.g. of Kratzer 1991). On the other hand, the stronger constraint makes slightly strange predictions about disjoined 'might's. Consider a context which contains \( p- \) and \( p \)-worlds, with \( g(i) \) an accessibility relation which is universal within the context. Then the stronger constraint predicts "Might, \( p \) or might, not \( p \)" is false throughout the context. This is prima facie wrong. On the other hand, we can get a true reading of "Might \( p \) or might, not \( p \)" under these assumptions by co-indexing the 'might's differently, so that we have e.g. "Might, \( p \) or might, not \( p \)", and then making \( g(j) \) the empty accessibility relation. This is a bit ad hoc, but shows that this consideration does not tell conclusively against the stronger constraint; in light of the behavior of modals under conditionals, I'm inclined towards the stronger constraint, but deciding between them is a task for another day.
(in this section I’ll abbreviate natural language sentences with a standard modal propositional language, just for readability):\(^{64}\)

(61) a. **Consecutive Update Idempotence**: For any sentence \( p \) with sentential constituents \( q \) and \( r \), if \([p]\) is defined anywhere in part in terms of the composite update function \([q][r] \), that composite update function must be idempotent.

b. **Idempotence**: \([q][r] \) is idempotent if and only if \( \forall c : c[q][r] = c[q][r][q][r] \)

This constraint says, essentially, that a composite update function is only defined if it is idempotent, where a function \( f \) is idempotent just in case, for any argument \( x \), applying \( f \) once to \( x \) yields the same result as applying it twice: \( f(x) = f(f(x)) \). This constraint rules out Wittgenstein disjunctions. The CCP of a Wittgenstein disjunction \( \Diamond(p \land -p) \lor (\Diamond q \land -q) \) is defined in part in terms of the composite update function \([\Diamond p][-p] \). But this function is not idempotent: consider a context \( s = \{w, w'\} \), with \( p \) true at \( w \) and false at \( w' \). Then \( s[\Diamond p][-p] = \{w'\} \), but \( s[\Diamond p][-p][\Diamond p][-p] = \emptyset.\(^{65}\)

So far, so good. But this constraint overgenerates. Consider a sentence like (62):

(62) a. John might be sick and Sue might be sick, but either John isn’t sick or Sue isn’t sick.

b. \((\Diamond p \land \Diamond q) \land (\neg p \lor \neg q)\).

(62) is perfectly felicitous. But it violates **Consecutive Update Idempotence**. The CCP of a sentence with the form of (62) will be defined in terms of the sequence \([\Diamond p \land \Diamond q][-p \lor -q] \). But this sequence is not idempotent. Consider a context \( s = \{w, w'\} \), with \( p \) and \( q \) both true at \( w \) and both false at \( w' \). Then \( s[\Diamond p \land \Diamond q][-p \lor -q] = \{w'\} \), but \( s[\Diamond p \land \Diamond q][-p \lor -q][\Diamond p \land \Diamond q][-p \lor -q] = \emptyset.\)**Consecutive Update Idempotence** thus wrongly predicts that sentences with the form of (62) will be infelicitous.

The bounded theory, by contrast, correctly predicts that sentences like (62) are perfectly consistent and coherent.

---

\(^{64}\)A similar constraint is given in Klinedinst and Rothschild 2014. That constraint does not predict that a sentence like (62) is uniformly infelicitous, but does predict that such a sentence will 'crash' a context like \( s \) to the empty set, which seems wrong.

\(^{65}\)And of course Wittgenstein disjunctions with right-embedded modals are already predicted to be incoherent by the standard update semantics.
If an approach along present lines is to solve our problems, then, we will have to find a principled way to restrict its application to avoid this kind of overgeneration. It is not clear to me how to do that.\textsuperscript{66}

1.8.2 Enriched update semantics (ii): Rothschild and Klinedinst 2015

A second, related enrichment to the update semantics is given in Rothschild and Klinedinst 2015, which also augments a standard update semantics with a further constraint:\textsuperscript{67}

(63) \textit{Successive Update Rule (SUR):} Let \([\cdot]\) be the standard update semantics denotation function. Let \([\cdot]\) be a new function from sentences to CCPs, defined recursively from \([\cdot]\) as follows: if there is no non-empty \(c\) such that \(c[p] = c[q] = c\), then \(c[p][q] = \emptyset\). In all other cases, \(c[\cdot] = c[\cdot]\).

This constraint is similar to Yalcin’s, but instead of stipulating that CCPs be defined only in idempotent terms, it instead stipulates that, unless there is some context which accepts each of two successive updates, then those successive updates yield the empty set. If we treat \([\cdot]\) rather than \([\cdot]\) as our semantic value function for natural language, then this approach captures the infelicity of Wittgenstein disjunctions. Those disjunctions contain the consecutive update \([\Diamond p][\neg p]\). No context accepts both \(\Diamond p\) and \(\neg p\). And so this update will take any context to \(\emptyset\); and so the disjunction as a whole will always crash any context to the empty set.

This approach avoids my objection to Yalcin’s idempotence rule, since, while \([\Diamond p \land \Diamond q][\neg p \lor \neg q]\) is not idempotent, there are contexts which accept both \(\Gamma (\Diamond p \land \Diamond q)\) and \(\Gamma (\neg p \lor \neg q)\) (e.g. a context \(s = \{w, w'\}\) with \(p\) true and \(q\) false at \(w\), and \(p\) false and \(q\) true at \(w'\)). Generally speaking, I think this approach is the most promising alternative account of Wittgenstein disjunctions. Let me bring out one observation, however, which I believe speaks in favor of my approach over this one: the \textit{SUR} is sensitive to syntactic structure in a way in which the bounded theory is not. To see this point,\textsuperscript{66} Yalcin himself floats the idea that the \textit{Consecutive Update Idempotence} rule applies only as a general preference, which is suspended when doing so would allow us to avoid disaster. If we took that tack, then we could explain the felicity of (62), but we would no longer explain the infelicity of Wittgenstein disjunctions, since presumably the constraint would just be suspended for left-modal Wittgenstein disjunctions, to avoid disaster. The difficulty is finding a principled way to explain why the constraint would apply to Wittgenstein disjunctions but not to (62).\textsuperscript{67} My presentation differs very slightly from Rothschild and Klinedinst (2015)’s.
compare (64) and (65), which differ only in where the negation is placed—a difference which does not generally effect the interpretation of sentences with this form:

(64)  
  a. Most people who might be sick aren’t sick.
  b. Most(might; p)(not p).

(65)  
  a. It’s not the case that most people who might be sick are sick.
  b. Not most(might; p)(p).

The SUR predicts a difference between sentences with the form of (64) and those with the form of (65). Given plausible semantic assumptions, the composition of (64) will go by way of a sequence with the form \([\Diamond p][-p]\) and thus will be predicted to be inconsistent by the SUR. By contrast, since negation takes high scope in (65), the composition of (65) will not go by way of any objectionable successive updates, and thus (65) is predicted to be consistent.

Contrary to these predictions, it seems to me that there is no contrast in felicity between (64) and (65). I have a slightly less clear judgment that both are felicitous. Suppose that 100 people tested positive for a given sickness, but we know that the test has a rate of false positives above .5. It seems to me that in this scenario I can felicitously assert (64-a) or (somewhat stiffly) (65-a). That both sentences have the same status, and that both are felicitous, is precisely the prediction of the bounded theory.

A similar point can be extended to other quantifiers, like ‘few’. Broadly speaking, the bounded theory predicts invariance under certain syntactic permutations which the SUR predicts will affect interpretation. A more thorough exploration may yet show that the latter claim is correct, but I believe the evidence surveyed here tells against it.

Let me close with a more abstract point against the SUR. There is a sense in which the SUR bleaches the asymmetry out of the update framework. But there is something strange about constructing an asymmetric framework in the first place, only to bleach the asymmetry out later on.

---

68See Rothschild and Klinedinst (2015) for a dynamic semantics for ‘most’ which bears this out.
69Sentences with the form of (64) and (65) will both be consistent on the bounded theory, since the local context for p in \("Most(p)(q)\) entails neither q nor \(\neg q\): \("Most(p)(q)\) is not in general equivalent to \("Most (p and q)(q)\) or to \("Most(p and not q)(q)\)."

---

56
Why not move directly to a symmetric framework? There may be high level arguments for this kind of approach, but, without such arguments, it looks unnecessarily roundabout.

1.8.3 Probabilistic contents: Moss 2015

The last system I'll discuss is due to Moss 2015. I will gloss over some complexity in Moss’s system which will be irrelevant for us, in particular concerning the role in her semantics for contextually salient partitions; I refer readers to the paper for formal details. The essentials are that every sentence ultimately denotes a set of probability measures. ‘Might’ takes as its complement a set of probability measures. Connectives come in two flavors: one which takes as arguments sets of probability measures, one which takes as arguments ordinary possible-worlds propositions. To let modal and non-modal sentences combine, and to ensure that every sentence ultimately denotes a set of probability measures, Moss introduces a type-shifter $C$ which takes a non-modal proposition $p$ to the set of probability measures built on sample spaces that entail $p$.

In a Wittgenstein sentence with the surface form $\neg \Box p \land \neg p \downarrow$, a modal sentence is conjoined with a non-modal one. We therefore need to insert type-shifters in order for this sentence be well-formed. Here are two possibilities:

\((66)\) $\Box C p \land C \neg p$.

\((67)\) $\Box C p \land \neg C p$.

These are both well-formed in Moss’s system. \((66)\) will say (very roughly) ‘Might $p$, and definitely not $p\downarrow$. In Moss’s system, no probability measure satisfies both these conjuncts (to do so, it would have to have $Pr(p) > 0$ and $Pr(\neg p) = 1$). And so \((66)\) will have as its semantic value the empty set. By contrast, \((67)\) says (very roughly) ‘Might $p$, and not definitely $p\downarrow$. There are probability measures that satisfy both conjuncts of \((67)\) (to do so, they only have to have $Pr(p) > 0$ and $Pr(p) < 1$).

If we stipulate that Wittgenstein sentences are always parsed as in \((66)\), rather than as in \((67)\) (for instance, by stipulating that type-shifters are always added as high as possible) then we will\(^70\) Or $\neg p \land \Diamond p \downarrow$; the points here go for either order.
predict that Wittgenstein sentences denote the empty set; so, too, will their disjunction.

So far, so good. But there is trouble in the neighborhood. Suppose we add a tautological modal conjunct to a Wittgenstein sentence, as in (68):\(^7\)

\[(68)\]

\[\begin{align*}
a. & \quad \#(p \wedge \neg (p \wedge 2+2=4)) \\
b. & \quad \#(p \wedge \neg (p \wedge \Box T))
\end{align*}\]

Clearly, adding a tautological conjunct to a Wittgenstein sentence in no way improves it. But adding a tautological conjunct has a curious effect in Moss's system: it forces a low type shifter. That is, (68) can only be interpreted as in (69); if we tried to put a 'high' type-shifter over the negation instead, the result would be the formula in (70), which illicitly mixes types in the right-most conjunction and is thus uninterpretable in Moss's system.

\[(69)\]

\[\diamond C p \wedge \neg (C p \wedge \Box C \top)\]

\[(70)\]

\[\diamond C p \wedge C \neg (p \wedge \Box C \top)\]

But (69) is semantically equivalent to (67), and thus will be interpreted, roughly, 'Might p and not definitely p'. This means that Moss's system predicts that adding a tautological modal conjunct to a Wittgenstein sentence, as in (68), will render it coherent; thus (68-a) is predicted to mean, roughly, 'It might rain, but it won't definitely rain'.

This, of course, is the wrong prediction. And this problem for Wittgenstein sentences can be immediately translated into a problem for Wittgenstein disjunctions: adding a tautological modal conjunct to Wittgenstein disjunctions is predicted to render them coherent. But this is plainly the wrong prediction.

There is, again, an element in Moss's system that I have ignored here for simplicity, namely a semantic role for partitions of logical space. It is true that, relative to some choices of partition, a sentence like (68) will still denote the empty set. A defender of Moss's view might try to argue that we always evaluate sentences like this relative to such partitions. But this is a difficult position to defend: what leads us to choose partitions which make us interpret speakers as expressing con-
tradictions, when there are other equally simple and relevant partitions in the neighborhood which render the speakers coherent? I do not see how to answer this question in a principled way; without an answer to that question, this approach does not provide a satisfying account of the infelicity of Wittgenstein sentences or Wittgenstein disjunctions.\textsuperscript{72,73}

1.9 Conclusion

There is strong reason to believe that "Might $p$" is classically consistent with "Not $p$"; otherwise "Might $p$" would entail $p$. And if two sentences are jointly consistent, it follows in classical logic that their conjunction is consistent. These facts have led nearly everyone studying epistemic modals to conclude that Wittgenstein sentences are classically consistent, and to ascribe their felt infelicity to something else—either to broadly pragmatic considerations, or to some kind of non-classical inconsistency.

I have argued that this approach makes a striking prediction: that the disjunction of two Wittgenstein sentences will itself be both consistent and assertable. While the corresponding prediction is borne out for disjoined Moore sentences, the prediction for Wittgenstein sentences is not.

I have tried to make sense of this situation by developing a new theory of epistemic modals, of embedding operators in natural language, and of their interaction. This theory builds on the standard, relational semantics for epistemic modals, but augments that theory with what I have called a locality constraint, which restricts the accessibility relations relative to which epistemic modals can be evaluated. The resulting approach is Strawson classical: classical entailments will

\textsuperscript{72}Thanks to Daniel Drucker for discussion on these points. Moss (2016), discussing a different but related problem, proposes that there is sometimes a background presumption that a relevant proposition either must be the case or can’t be the case. Adding this presumption would, of course, strengthen these problematic sentences to contradictions. But there is no reason to think that speakers would make this assumption when evaluating a sentence like (68-a), since this assumption is straightforwardly inconsistent with the asserted content (and adding this assumption as a non-defeasible background presupposition is, of course, a non-starter, since that would render incoherent ordinary sentences like ‘It might be raining, and it might not be raining’).

\textsuperscript{73}A related system to Moss’s, developed by Hawke and Steinert-Threlkeld (2016), avoids this issue by doing away with type-shifting. On this system, every sentence ends up denoting a set of information states, but the system is built on information states all the way down, and thus has no need for type-shifting. This semantics predicts that Wittgenstein disjunctions are inconsistent, and it avoids Moss’s problem with tautological conjuncts. This semantics has serious problems with negation, however: it predicts that "Not must $p$" is equivalent to "Not $p". But see Aloni 2016, Steinert-Threlkeld 2017, Chapter 3 for recent bilateral developments which avoid this problem; these systems look to me like a promising way to account for these data in a similar framework to the present one.
be valid, provided they preserve definedness. Crucially, this caveat lets us block the argument from the consistency of "Might p" with "Not p", to the consistency of "Might p and not p" (or order variants); the latter will be undefined, relative to any variable assignment and world which makes both of the former true at any world. Thus "Might p" will be classically consistent with "Not p", but their conjunction will not be classically consistent. This provides an account of the infelicity of Wittgenstein sentences which, crucially, extends immediately to an explanation of the infelicity of Wittgenstein disjunctions, and thus of their divergence from Moore disjunctions.

I have argued, moreover, that Wittgenstein disjunctions show that the dependence of epistemic modals depend on the intrasentential dynamics of information is fundamentally symmetric. This fact, in turn, has striking ramifications for the semantics of epistemic modals. I have argued that it is not possible to capture these symmetric dynamics in a satisfactory way in a domain or update framework; the relationality of my theory is fundamental to its account of Wittgenstein sentences. That, in turn, shows that—contrary to much recent work—it is not only possible to account for the behavior of epistemic modals in a broadly relational framework, but we in fact need the resources of that framework to make sense of the embedding behavior of epistemic modals.

The bounded theory is logically conservative, in two senses. First, again, it (nearly) preserves classical inference rules for the logical connectives, insofar as the resulting system is Strawson classical. Second, unlike most other attempts to account for the embedding behavior of epistemic modals (like the update or domain semantics), the bounded theory is non-committal about key principles in modal logic like positive or negative introspection. Whether or not those principles are valid is not my concern here; my point is just that we can make sense of the embedding behavior of epistemic modals which we have dealt with here without taking a stand on those principles.

I have shown, finally, that the bounded theory makes strikingly accurate predictions, not only about Wittgenstein disjunctions, but also about the embedding behavior of epistemic modals in a wide variety of other embedding environments. And, importantly, the bounded theory makes these predictions in a uniform and principled way: given a classical semantics for an embedding operator, the bounded theory makes a prediction about how modals will behave under that operator, without having to make ad hoc stipulations about how embedding operators shift the interpretation.
of epistemic modals. This suggests to me that the bounded theory of epistemic modality is not only an advance in extensional adequacy, but also in explanatory accuracy, in the theory of epistemic modals.

In concluding, let me comment briefly on the two parts of the bounded theory. First, the bounded theory makes essential use of the theory of local contexts developed in Schlenker 2009. The ability of that algorithm to capture the subtle data discussed here speaks substantially in favor of the claim that it captures something fundamental about natural language. But the present implementation of Schlenker's algorithm departs from the way Schlenker envisioned it in a few important ways. One involves order, as discussed above. Another concerns the fact that the bounded theory assumes that local contexts are semantically accessible. This is a departure from the pragmatic way in which Schlenker (2008a) introduced his algorithm in (building on earlier pragmatic work in Stalnaker 1974; Karttunen 1974). Both these facts may have important upshots for the theory of presupposition and, generally speaking, the dynamics of information in natural language.

Second, if the bounded theory is correct, it, again, vindicates a broadly relational approach to epistemic modals. Let me note two upshots of this. First, insofar as expressivist and relativist approaches have generally been spelled out in quite different semantic frameworks, the vindication of a broadly relational approach may have an indirect bearing on the debate between contextualists, expressivists, and relativists, about how accessibility relations are chosen (the locality constraint only puts a bound on the choice of accessibility relations; it does not fully specify what factors determine the complete choice of an accessibility relation). It seems open to me that a relativist or expressivist interpretation of the bounded theory could be given, but it strikes me that it would be a somewhat awkward marriage. Second, the bounded theory semantics to some degree leaves intact the Kratzerian promise of a unified theory of modality across modal flavors. Having said that, not all modals are subject to the locality constraint characteristic of epistemic modals; while probability modals like 'probably' and 'likely' certainly are (a fact which can easily be captured by building semantics for those modals on top of the architecture of the bounded theory), root modals (deontic, agentive, and circumstantial), as well as weak necessity modals ('ought', 'should') do not seem
to be subject to a similar constraint.\footnote{Such a constraint, interestingly, would explain some of the puzzling data for deontic ‘must’ in Ninan 2005, but doesn’t seem tenable for deontic ‘may’.} This still leaves open a broadly unified approach to natural language modals, though: the idea would be that what semantically distinguishes epistemic from root modality just is the locality constraint. This idea requires much further exploration.

The bounded theory of epistemic modality shows that there is something right in two of the main approaches to epistemic modals. On the one hand, as the standard relational view has it, the truth conditions of epistemic modals are close to those of modal operators in modal logic: ‘Might $p$’ is true only if $p$ is true in some accessible world. On the other hand, as dynamic semanticists have argued, the truth conditions of epistemic modals are sensitive to what information is locally available when we process them—though in a way which turns out to be starkly at odds with the dynamic approach to epistemic modals. Thus, while the core meanings of ‘Might $p$’ and ‘For all we know, $p$’ turn out to be closely related, the two interact with their local information in very different ways: interpretation of the former, but not the latter, is constrained by its local informational environment. Why are epistemic modals constrained in this way, and what is the connection to presupposition, which seems to be dependent on its local information in a similar way? These are big questions which require much further inquiry. I suspect the answer will build on the insight that epistemic modals are used to coordinate on information in conversation; the locality constraint makes epistemic modals a remarkably subtle tool for coordinating information in a local, and not just global, fashion.
Chapter 2

How to do things with modals

In this chapter, I turn from the question of what epistemic modals mean to the question of how speakers use those meanings to coordinate on their common information. I show that—pace much recent work—we can account for the core dynamics of epistemic modality within a broadly relational semantic framework together with a standard contextualist approach to the dynamics of conversation.

2.1 Introduction

In a brief but trenchant discussion of epistemic modals, Wittgenstein warns against ‘regard[ing] a hesitant assertion as an assertion of hesitancy’.¹ An epistemic modal claim like ‘It might be raining’, the thought goes, should not be regarded as an assertion of the speaker’s uncertainty as to whether or not it is raining, but rather as something quite different in kind: a proposal to treat the possibility of rain as live.

Wittgenstein’s admonition provides a helpful lens for viewing the subsequent debate about the meaning of epistemic modal claims, and the system of communication that these claims fit into. On the one hand, relational accounts (including the non-standard relational account presented in Chapter 1) treat ‘It might be raining’ as expressing an ordinary piece of information, namely that

¹ ‘Betrachte nicht die zaghafte Behauptung als Behauptung der Zaghaftigkeit’ (Wittgenstein, 1953, II.x.110).
the proposition that it’s raining is compatible with some contextually salient piece of evidence. On the face of it, relational accounts thus treat modal claims as assertions of hesitancy, in Wittgenstein’s phrase; because of this, they have been taken to task for failing to make sense of how speakers do what they do with modal claims.2 On the other hand, an array of heterodox accounts have taken Wittgenstein’s admonition to heart, re-engineering not only the standard theory of the meaning of epistemic modals, but also the contextualist theory of communication—on which communication is the transfer of information—in order to make sense of the dynamics of modal language.

In this chapter I will show that there is a way to walk a line between these two options. It is indeed wrong to view a ‘might’-claim as an assertion of uncertainty, as a flat-footed interpretation of the relational theory would have it. This approach fails to capture a fundamental observation about the conversational dynamics of epistemic modality, namely, that an assertion of ‘Might p’ is a proposal to make p compatible with the common ground;3 a proposal which, if accepted, ensures that p is compatible with the common ground, and, if rejected, ensures that it is not. I call this observation about the fundamental dynamics of epistemic modality the guiding observation.

It is overhasty, however, to conclude—as many in the recent literature have—that there is no content that we can assign to ‘might’-claims which will capture the guiding observation. I show, to the contrary, that we can capture the guiding observation, and therefore the fundamental dynamics of epistemic modality, within the contextualist framework. We can do so by assigning an assertion of ‘Might p’ a content which is about the conversation’s common ground itself. In particular, we maintain that ‘Might p’ says that p is compatible with the common ground after the claim in question has been made and negotiated. I show that this theory of the assertoric content of modal claims (which I call prospective contextualism), together with the contextualist theory of communication, guarantees that an assertion of ‘Might p’ just is a proposal to make p compatible with the common ground, thus capturing the guiding observation.

Prospective contextualism thus shows that we can make sense of the fundamental dynamics of epistemic modality within the contextualist framework. The main goal of this chapter is to show just

---

2I will often use ‘modal claim’ for ‘epistemic modal claim’; my focus here, as throughout the dissertation, will be exclusively on modal claims interpreted in a broadly epistemic sense.

3The set of propositions commonly accepted in the conversation.
this: that, *pace* much recent work, a contextualist approach to epistemic modality cannot be rejected on the putative basis that it cannot capture the guiding observation. Whether prospective contextualism is the *right* way to model the dynamics of epistemic modality is another, much broader question. I will not try to answer this question decisively here. I will try to say enough, however, to show that prospective contextualism should be taken seriously as a contender theory. In particular, I will argue that some natural objections to prospective contextualism can be answered, by bringing out a parallel between the structure of modal claims, on my theory, and the structure of *performative* uses of language; and that prospective contextualism has some at least *prima facie* advantages over some competitor views, to a degree which makes it worth serious further exploration. I conclude by suggesting that the present approach to the dynamics of modality can be extended to give a contextualist model for a wide variety of other speech acts which at first blush appear to go beyond the bounds of contextualism.

### 2.2 Contextualism and its discontents

I begin by sketching the contextualist framework, and the challenge that epistemic modal claims pose to it.

#### 2.2.1 Contextualism

‘Contextualist’ is sometimes used to describe a theory of how a particular term works; for instance, any plausible theory of the meaning of ‘I’ will be contextualist, insofar as it will say that the denotation of ‘I’ varies with the context (it always denotes the speaker of the context). I will use ‘contextualist’ here in a more general way, to describe a wholesale model of communication and division of labor between semantics and pragmatics.\(^4\) In the contextualist framework, semantics is responsible for assigning contents to well-formed declarative sentences, relative to contexts;\(^5\) this process will be compositional, and the output of our linguistic cognitive system. The contents which

---


\(^5\)More precisely, to logical forms. A *context* provides information about anaphora resolution, etc.; standardly, we can think of a context just as a variable assignment.
are the output of this process of interpretation will be units of information, or propositions, which we model as sets of possible worlds—intuitively, the set of all and only those worlds which verify the proposition in question.⁶

Pragmatics, by contrast, concerns the application of domain-general reasoning to linguistic communication. Most importantly for our purposes, the propositions which are the output of the interpretation of sentences at contexts will interface with a pragmatic system, which does things with these contents. In particular, contextualism commits to a simple theory of the basic thing speakers do with these contents: they assert them in order to coordinate on their information. To model this, contextualism tracks the common commitments in a conversation at a given time—the conversation’s common ground—and then says that an assertion of a declarative sentence is a proposal to add the proposition expressed by the sentence at that context to the common ground.⁷ An assertion, in other words, is just a proposal for the interlocutors to come to commonly believe the content asserted.

I make all this more formal in Appendix B, but this informal exposition suffices for present purposes. Contextualism provides an elegant model of the division of labor between semantics and pragmatics, and an elegant model of the dynamics of conversation. And, since it views conversation as the mutual exchange of information, contextualism provides a fully explicit model of the evolution of conversation (at least once we couple it with an independently motivated theory of how agents assimilate new information).

2.2.2 Discontents

Despite its attractions, however, it is easy to get in a mood in which contextualism looks too simple. Conversation is replete with uses of language which, at first blush, go far beyond the simple

⁶This latter modeling choice is strictly independent of the contextualist framework; all that is really crucial is that propositions can be regarded as units of information.
⁷More explicitly, the common ground is the set of propositions accepted by all conversants, accepted to be accepted, and so on, where acceptance is an attitude like belief or knowledge. I will often conflate acceptance and belief in what follows for the sake of concreteness. Note that the common ground is defined for any number of conversants, including just one (in that case the common ground will be identical to the transitive closure of the individual’s state of acceptance). A natural objection to the view I will give below is that it cannot make sense of what we are doing when we use modal language when thinking or talking to ourselves, since it references the common ground; but this objection is premised on the mistaken assumption that the common ground is an essentially interpersonal notion.
exchange of information. In the rest of this chapter I will make a case study of one use of language which poses this problem in a particularly distinct and challenging way: epistemic modal claims.

What do speakers do with epistemic modal claims? Focus for now on ‘might’-claims, like ‘it might be raining’ or ‘John might be in his office’. Intuitively, a ‘might’-claim is a way of proposing that some possibility should be taken seriously in inquiry: e.g. ‘It might be raining’ says ‘Let’s take seriously the possibility that it’s raining’. In the contextualist framework, we can make this precise by saying that a claim of ‘Might \( p \)’ is a proposal to make \( p \) compatible with the common ground, and to make this fact itself common ground. Thus an assertion of ‘John might be in his office’ is a proposal to make it compatible with the common ground that John’s in his office, and to make this fact itself common ground.

This description of the dynamics of modality, which I will refer to in what follows as the guiding observation, makes sense of intuitions about what speakers are trying to do when they assert ‘might’-claims: namely, to ensure that we take seriously its prejacent (the proposition it embeds) as a live possibility in our investigation of what the world is like. It also makes sense of intuitions about what speakers are doing when they agree or disagree about ‘might’-claims. If you say that John might be in his office and I object that he can’t be, what we are arguing about is whether to treat as live the possibility that John is in his office: you are saying we should, I am saying we should not. If we treat ‘must’ as the dual of ‘might’, then the guiding observation also makes sense of the dynamics of ‘must’-claims, predicting that an assertion of ‘Must \( p \)’ is a proposal to ensure that \( p \) is entailed by the common ground. This, again, conforms to intuitions and makes sense of agreement and disagreement about ‘must’-claims, though this will not be our focus here.

Now that we have the guiding observation in hand, we can formulate the challenge epistemic modals pose to contextualism. The issue is that an assertion, on the contextualist theory, is a proposal to accept some piece of information. But, if the guiding observation is right, a ‘might’-claim does not look like a proposal to accept a piece of information, but rather a proposal to make some piece of information compatible with the common ground: in Wittgenstein’s phrase, a hesitant assertion,

---

67

---
not an assertion of hesitancy.

To make this worry more pressing, consider the standard relational theory of the meaning of epistemic modal claims. On this theory, 'Might p' means that p is compatible with a body of evidence or attitude state made salient by the context of utterance.\(^9\) This is modeled by associating ‘might’ with a contextually provided accessibility relation between worlds, on which w accesses w' just in case w' is compatible with the contextually salient evidence or attitudes at w. Then 'Might p' is true at context c and world w just in case p is true at some world accessible from w.

At first blush this looks reasonable enough, but, when we try to say more about how the accessibility relation in question is actually determined at given contexts, it is hard to see how to bring this theory in line with the guiding observation.\(^10\) A natural first thought would be to take the accessibility relation to be determined by the knowledge of the speaker. Then 'Might p' will be equivalent to 'For all I know, p'. But this clearly doesn't conform to the guiding observation: a claim with the form 'For all I know, p' is not a proposal to treat p as compatible with the common ground, but a statement that p already is compatible with what you know (and, therefore, with the common ground). This model thus fails to account for how speakers negotiate about modal claims. If you say ‘John might be in his office’ and I reply, ‘No, he’s in England’, I am not disputing that it is compatible with your antecedent knowledge that John is in his office (after all, you should know); instead, I’m disputing whether this is a good thing for us to leave open in our inquiry. Likewise, if I agree, I am not agreeing that it is compatible with your knowledge that John is in his office (again, you should know), but rather that it’s a good thing to leave open in inquiry.

A natural second attempt would take the accessibility relation to be determined, not by the speaker’s knowledge, but rather the knowledge of the group of conversants: the knowledge state that would result if all the individuals in the group pooled their knowledge.\(^11\) But this approach still fails to conform to the guiding observation. Suppose we’re in a big group. You see John’s office

---


\(^11\)In standard terminology, the group’s distributed knowledge: i.e., where \(K_{S,w}\) represents the set of worlds compatible with S’s knowledge in w, and I is a group, \(\bigcap_{i \in I} K_{i,w}\). A weaker approach, which takes the accessibility relation to be determined by the weakest thing known—i.e., by \(\bigcup_{i \in I} K_{i,w}\)—would fail to make sense of disagreement.
light on. I ask you where John is. You can say, ‘He might be in his office’: you are proposing to treat as live the possibility that John is in his office. You are not asserting that, if everyone in the group pooled their knowledge, it would remain compatible with our knowledge that John is in his office. Indeed, you may think that, if we pooled all our knowledge, we might well discover that John was not in his office. More succinctly: in cases like this, ‘John might be in his office’ is intuitively assertable, whereas ‘The strongest thing we together know leaves open that John is in his office’ is, intuitively, not. These thus cannot have the same content in this context.

These two options do not exhaust the space of possibilities—a point I’ll return to in the next section—but it is tempting at this point to echo Wittgenstein: the standard theory, and the contextualist framework in which it is embedded, makes the mistake of trying to analyze modal assertions—a special kind of speech act whose aim is to coordinate on which possibilities to treat as live—as assertions of hesitancy. It is a mistake to analyze epistemic modals as any kind of assertion at all, the thought goes; modal claims do not aim to transfer information, but rather constitute a different speech act, something which takes us altogether out of the bounds of the contextualist framework. This pessimistic line is implicit or explicit in much of the recent work on epistemic modals; here is a characteristic statement:12

In general, there is no single proposition one can accept such that one doesn’t rule out $p$ if and only if one accepts that proposition ($p$ itself is too strong). So if our language provides the resources to simply express that we don’t rule out $p$, then it seems we will not be able to understand that bit of language if we confine ourselves to a framework where sentences express propositions. (Rothschild, 2013, p. 50)

\[12\] I should note that Rothschild (2013) is not clear about whether he endorses this view. Another characteristic statement comes in Swanson (2011, p. 251): ‘Construing subjective uncertainty about whether $\varphi$ in terms of near certainty about some other proposition seems wrongheaded. But unless the truth conditional theorist can find such propositions, there is no reason to suppose that an assertion of a doxastically hedged sentence will inculcate the appropriate partial belief in the addressee. And the project of finding such propositions looks quixotic if not impossible.’ Accounts which follow this pessimistic line in rejecting the contextualist framework to capture the guiding observation include dynamic accounts such as Veltman 1996; Beaver 2001; Willer 2013; expressivist accounts such as Yalcin 2007, 2012a, 2011; Rothschild 2011; Swanson 2015; Moss 2015; revisionary contextualist accounts such as von Fintel and Gillies 2011; Stalnaker 2014, and relativist accounts such as Egan et al. 2005; Stephenson 2007b; Köhl 2009; Egan 2011; MacFarlane 2011, 2014. I discuss some of these approaches in §§2.5 and 2.7 below. A different defense of contextualism from the one I give below comes from ‘flexible’ contextualist accounts such as Dowell 2011; Khoo 2015, as well as the coherence theory version of contextualism in Stojnić 2016; for reasons of space, I cannot discuss those interesting views here.
2.3 Prospective contextualism

But this pessimistic conclusion is too fast. *Pace* this pessimistic conclusion, there *is* a proposition we can assign as the content of "Might p" which captures the guiding observation. The basic idea is to preserve the structure of the standard theory, but to have the accessibility relation determined by *the common ground itself*. In this section I show that, if we do this in the right way, then, thanks to certain properties of the logic of common ground, an assertion of "Might p" will amount to a proposal to make p compatible with the common ground (and to make this fact itself common ground).

As a first pass at implementing this idea, we could take 'It might be raining' to mean, roughly: the proposition that it's raining is compatible with the common ground. But this wouldn't work. On this view, modal claims would merely *describe* the common ground. Like the views discussed in the last section, this cannot make sense of the way that speakers use 'might'-claims to *negotiate* about the common ground. If I say that it might be raining, and you say it is not, we are not arguing about whether the common ground is *presently* compatible with the proposition that it’s raining. We know this in part because if you think that the common ground is compatible with p, then it follows from the logic of the common ground that the common ground *is* compatible with p. We are, rather, arguing about whether the common ground *should be* compatible with the proposition that it’s raining. A parallel point extends to 'must'. On the present account, "Must p" would say that p is already entailed by the common ground. That would mean that ‘It must be raining’ could not be used to *inform* one’s interlocutors that it’s raining; it could only be truly asserted in a context where it is already accepted that it’s raining. This again fails to capture the way in which ‘must’-claims are used to negotiate about the common ground.

A close variant on this approach, however, avoids these problems. On this view—which I call *prospective contextualism*—'It might be raining' means that the proposition that it's raining *will*
be compatible with the prospective common ground: the common ground as it stands after the assertion is made and negotiated, i.e. after the assertion has been either accepted or rejected by all interlocutors. According to prospective contextualism, ‘Might p’ will mean that p is compatible with the prospective common ground, ‘Must p’ that p is entailed by the prospective common ground.

On this theory, an assertion of ‘Might p’ is just a proposal to make p compatible with the common ground (and to make this fact itself common ground). Here’s why. Suppose that ‘Might p’ is asserted at time t; let t’ be the time when the assertion has just been either accepted or rejected. Suppose first that ‘Might p’ is accepted. In the contextualist framework, that means, again, that its content will be added to the common ground. Thus, at t’, the proposition expressed by ‘Might p’ will be in the common ground. If ‘Might p’ has the meaning attributed to it by prospective contextualism, it follows that it will be common ground at t’ that the common ground at t’ is compatible with p. This is not yet what we want, but it turns out that in the logic of common ground, ‘It is common ground that p is compatible with the common ground’ entails ‘The common ground is compatible with p’. In other words, when the common ground thinks something is compatible with it, it is always right. Importantly, this does not follow from any (non-trivial) assumptions about the logic of the attitudes which constitute the common ground, but rather from the iterative structure of the common ground, on which the common ground is what is accepted, accepted to be accepted, and so on. Given this fact about the logic of the common ground, it thus follows that an assertion of ‘Might p’ is a proposal which, if accepted, ensures that p is compatible with the common ground (at the time after the assertion has been negotiated), and that this fact itself is common ground.

17What the prospective common ground actually amounts to is of course a vague matter; in this respect the notion is on a par with standard notions in semantics and pragmatics, like the notion of a context. Importantly, the prospective common ground differs from the common ground at the point just after it has been updated with the fact that the claim has been made (see Stalnaker 1998; von Fintel 2008. Quantifying over that common ground does not avoid the problems just discussed for the non-prospective view.

16Again, models for common ground for a group I can be constructed by taking the transitive closure of the accessibility relations for acceptance for each i ∈ I, where the transitive closure of a set of relations R: i ∈ I is R*, where xR* y just in case there is a sequence of worlds w_1...w_n such that w_1 = x, w_n = y, and, for each w_i and w_{i+1} for i < n - 1, ∃k ∈ I : w_i R_k w_{i+1}. Then, interpreting □_i as ‘it is common ground at t that’, and ◊_i as its dual, this claim is just the claim that □_i ◊_i p → ◊_i p is a theorem in the logic of common ground. To prove this, we need only to assume that the underlying attitude of acceptance is always consistent, in which case the system validates □_i p → ◊_i p (corresponding to the serial constraint). Substituting ◊_i q for p, we get ⊢ □_i ◊_i q → ◊_i ◊_i q. Any system for common acceptance will validate □_i p → □_i □_i p, thanks to its iterative structure (corresponding to transitivity); contraposing and substituting q for ¬p, we get ⊢ ◊_i ◊_i q → ◊_i q. It follows that ⊢ □_i ◊_i q → ◊_i q.
Suppose second that ‘Might p’ is rejected by all parties to the conversation; in the strong sense of rejection I have in mind, it follows that its negation is common ground at t’. According to prospective contextualism, the negation of ‘Might p’ says that the prospective common ground, i.e. the common ground at t’, is not compatible with p; that is, that it entails \( \overline{p} \). Thus it will be common ground at t’ that \( \overline{p} \) is common ground. This, again, is not yet what we want, but, once more, the logic of common ground helps us close the gap. It follows in the logic of common ground that, if the common ground entails that the common ground entails some p, then the common ground entails p. In short, when the common ground thinks that it entails something, it is always right.\(^7\) Thus if ‘Might p’ is commonly rejected at t’, then, at t’, \( \overline{p} \) will be common ground.

In sum: an assertion of ‘Might p’, according to prospective contextualism, will amount to a proposal which, if commonly accepted, makes p compatible with the common ground (and makes this fact itself common ground); and which, if commonly rejected, makes \( \overline{p} \) common ground. If we define ‘must’ as the dual of ‘might’, exactly parallel reasoning shows that ‘Must p’ is a proposal which, if accepted, adds p to the common ground; and, if rejected, ensures that p is not common ground, and thus that \( \overline{p} \) is compatible with the common ground.\(^8\)

Prospective contextualism, together with the contextualist model of conversation dynamics, thus predicts that assertions of ‘Might p’ and ‘Must p’ will have exactly the update properties ascribed to them by the guiding observation. In this framework, when we accept ‘Might p’, we accept a proposition which says that p is compatible with our own (prospective) common ground. And to accept such a proposition just is to render p compatible with our common ground. Likewise, when we accept ‘Must p’, we accept a proposition which says that p is entailed by our own (prospective) common ground. And to accept such a proposition just is to make p entailed by our common ground. Prospective contextualism thus shows that the guiding observation can be reconciled with

\(^7\)This amounts to the claim that \( \square \overline{\square}p \rightarrow \square \overline{p} \) is a theorem in the logic of common ground. Known as C4, this corresponds to the density constraint on accessibility relations (\( \forall a \forall b : a R b \rightarrow (\exists c : a R c \land c R b) \)), but it is perhaps more intuitive to derive it from the quasi-reflexivity constraint (every world which is accessed by a world accesses itself, corresponding to the constraint that attitudes of acceptance represent themselves as being veridical). Then the logic of common ground will validate \( \square \overline{\square}p \rightarrow \square \overline{p} \). By the K axiom and modus ponens we have \( \vdash \square \overline{\square}p \rightarrow \square \overline{p} \).

\(^8\)An assertion of ‘Must p’ is thus predicted to have the same main update effect as an assertion of p alone. As von Fintel and Gillies (2010) argue, this is broadly plausible, modulo some subtle but important differences; see Chapter 3 for an account of these differences within the present framework. One issue which von Fintel and Gilles make much of is whether ‘Must p’ entails p. On my account, the answer to this question will depend on whether we treat the common ground as factive or non-factive. It does not matter for the purposes of this chapter which way we go.
the contextualist framework: we can view modal assertions as proposals to add a proposition to the common ground, and still make sense of the fundamental dynamics of modal language, provided that we take modal claims to express propositions which are themselves about the common ground.

Before moving on, let me note two features of prospective contextualism. First, prospective contextualism is a theory about the assertoric content of unembedded modal claims: what proposition they contribute to the common ground when asserted. In §2.6, I’ll say a little bit about how to implement this theory compositionally, and in Appendix B, I will formalize the main claims of this section. But my main concern in this chapter is not with how to semantically generate prospective contextualism—and, therefore, not with the question of how modals embed—but rather how this theory of the assertoric content of modal claims answers to the guiding observation.

Second, what is essential to prospective contextualism is that it assigns to a modal claim a content which has certain introspective properties with regard to the common ground: in particular, where \( t \) is the time of assertion and \( t' \) is the prospective time, and annotating our modal claims with the time of assertion, it has the properties that \( \Box t' (\text{Might}_{t} p) \) entails \( \Diamond t' p \); and that \( \Box t' (\text{Must}_{t} p) \) entails \( \Box t' p \). Making epistemic modals about the prospective common ground is one way of accomplishing this, but there may be other ways—for instance, by making them about the prospective common evidence, and adopting plausible introspective constraints regarding evidence. This is a route that I will briefly explore in the next chapter.

2.4 Performativity

Prospective contextualism shows that, pace much recent work, we can capture the guiding observation within the contextualist framework. It does not follow, of course, that we should do so. The main goal of this chapter is simply to show that contextualism can capture the guiding observation, and thus that we can dismiss any rejection of contextualism based on its putative inability to do so. Whether prospective contextualism provides the correct theory of epistemic modality is not something I will take a stand on here; this question turns on a wide array of issues which I cannot hope to address in this chapter. In the rest of the chapter, however, I will try to dispel some obvious concerns
about prospective contextualism, and bring out some prima facie attractions vis-à-vis some of its competitors; I hope to say enough, not to show that prospective contextualism is the correct theory of epistemic modality, but rather to show that we should take prospective contextualism seriously as a contender theory.

I will begin by addressing a natural worry about prospective contextualism, which goes like this: There is something very weird about prospective contextualism. It captures the guiding observation in a way which is both structurally and normatively implausible. According to prospective contextualism, an assertion of an epistemic modal claim is a proposal about how the common ground should evolve which is made by saying something about how it will evolve. This is roundabout: the truth conditions in this account are idle wheels. And this account is normatively implausible: speakers can assert modal claims without having any idea of what will actually happen to the common ground, and thus without knowing or believing the content which prospective contextualism assigns to them. Given that you generally have to believe what you say, prospective contextualism can't be the right theory of epistemic modals.

I will address this worry by showing that the method of negotiation which I have argued epistemic modals exploit—making a proposal about how some contextual parameter ought to be set by making a truth conditional claim about how it will be set—is in fact widespread in natural language. In particular, epistemic modal claims, in this approach, recapitulate the structure of a wide variety of performative assertions. This fact shows that there is nothing suspicious about the structure which prospective contextualism attributes to epistemic modal claims. It also puts us in a position to address the normative issue for prospective contextualism, by arguing that performatives are governed by norms which do not require one to know or believe their content—explaining how speakers can assert modal claims, despite not believing the content ascribed to them by prospective contextualism.

19 This treatment of epistemic modals as performatives can be seen as an extension of the approach in Lewis 1979a, which treats some uses of deontic modals—syntactic and semantic siblings of epistemic modals—in this way.
2.4.1 Performatives in natural language

Suppose Mark tells his son John ‘This afternoon, you’ll be cleaning the rabbit cage.’ Mark’s claim has straightforward truth-conditions: it’s true just in case John will be cleaning the rabbit cage this afternoon. But his aim in asserting it is not (just) to inform John about the future, but to make that future come about: to establish certain normative facts about what John may do, and to do so by making a claim about what he will do. To get some more examples of *performatives* like this, consider a doctor telling a patient, in the presence of a nurse, ‘The nurse will now take you to the operating theater’ (Anscombe, 1963); or a sign at the front of a train track, which says ‘The front of the train will stop here’ (Robert Stalnaker, p.c.). These examples attest exactly the same performative structure that I am attributing to epistemic modal claims. It is clear that these sentences have ordinary truth-conditions; their performativity is not semantically encoded in any sense. One way to see this is that a sentence like ‘The nurse will now take you to the operating theater’ can be used *both* to inform the patient of what will happen next, and to give the nurse an instruction. So these sentences simply make claims with ordinary truth conditions. But these assertions make proposals about how to change a feature of the context (in these cases, normative features) by making a claim with ordinary, but prospective, truth conditions; a claim which, if accepted, ensures that the change in question takes place.

To further bring out the parallel between modal claims and performatives, let me draw attention to a fascinating analogy between the two. Performative language generally has non-performative usage.²⁰ For instance, Mark may very well say ‘You’ll be cleaning the rabbit cage’ just as a way to describe what John will do, *not* as a way to tell John what to do. Modal claims have exactly the same kind of non-performative uses, brought out in ‘stand your ground’ cases discussed in von Fintel and Gillies 2011. If Ann says ‘The keys might be in the car’ and Bill responds that they’re not, Ann might reply: ‘Look, I didn’t say they were in the car. I said they might be there—and they might have been.’ Here, we can interpret Ann as insisting that her initial assertion in was not a performativ but rather a description of the common ground at the time of the assertion (at that time, it was compatible with the common ground that the keys were in the car, since for all Ann

---

²⁰See Lewis 1979a for this observation regarding deontic modals.
knew, they were). Modal claims thus have exactly the duality that most performatives have: they are typically used to negotiate about the common ground, but can also, in the right circumstances, be used in a non-performative way, simply to describe the common ground. (I discuss further how to capture this fact semantically in §2.6.)

This discussion shows that the performative structure that prospective contextualism attributes to modal claims is, in fact, widely attested in natural language.\footnote{See Truckenbrodt 2009, Eckardt and Schwager 2009, Eckardt 2009 and Condoravdi and Lauer 2011 for similar ideas about performatives.} This should allay fears that that structure is implausibly roundabout.\footnote{It may be objected that these cases are explicitly marked out as being about the future, whereas epistemic modal claims are not. This is true. But note that prospective contextualism does not claim that modal claims are interpreted \textit{in the future tense} (a claim which would be farfetched), but simply that they are assigned accessibility relations in a way which references the future.}

### 2.4.2 Norms for modals: The perspective from performatives

It also helps us answer a natural objection concerning norms of assertion. The concern, again, is that, if modal claims have the content I am ascribing to them, then speakers will often be able to assert epistemic modal claims without knowing, or even believing, that their content is true. I can propose, for instance, to make the common ground entail that Sue is in her office, by asserting ‘Sue must be in her office’, without knowing or even believing that the common ground will come to entail this: for all I know, you may fail to accept my proposal. If epistemic modal claims were governed by a norm along the lines ‘Only assert what you believe (or know)’, prospective contextualism would leave it puzzling how people could assert epistemic modal claims in ordinary circumstances. I will argue in this section that a similar puzzle arises for performatives in general, which shows that performative assertions are not governed by a belief or knowledge norm of this form in the first place, but rather by a norm oriented towards what agents do with their words. This provides a principled explanation of how, and when, we can reasonably assert epistemic modal claims even if we do not know or believe their contents.

Let us consider first performatives like those considered in the last subsection, such as ‘This afternoon, you’ll be cleaning the rabbit cage’. Could a performative like this be governed by a norm
along the lines: Assert p only if you know (or ‘believe’, or ‘justifiably believe’) p?\(^{23}\) Clearly not. Mark may well tell John that he’ll be cleaning the rabbit cage as an attempt to get him to do it, even though he knows this attempt may well not be successful. One way to see this point is to compare a sentence like ‘You’ll be cleaning the rabbit cage’ to ‘You have to clean the rabbit cage’. In many contexts, these sentences play the same role: to try to get John to clean the rabbit cage. But it would be ludicrous to think that the latter of these is accompanied by a norm which holds that it can be asserted only if the speaker knows the addressee will clean the rabbit cage. And it seems no more plausible to hold that such a norm governs the former. This, of course, is in stark contrast to non-performative assertions: if Mark asks John what he was doing in the morning, then it is generally not permissible for John to say that he was playing outside, if he doesn’t believe this to be true.\(^{24}\)

This observations are not hard to make sense of. Performatives are used to do things, and so the norms that govern their production (and negotiation) will concern in the first instance what the performative is being used to do, and only derivatively about the speaker’s doxastic relation to its content. In short, performatives are governed by a very general norm along the lines: Assert p only if the action which you aim to accomplish with your assertion is permissible. Call this the *Speech Act Norm*. This norm rightly predicts that whether or not Mark may tell John that he will be cleaning the rabbit cage depends on whether it is permissible for him to require John to clean the rabbit cage—not whether he knows, or believes, that John will clean the rabbit cage.

All assertions, of course, are speech acts. As such, I propose that the *Speech Act Norm* governs assertions in general, not just performative assertions. The *Speech Act Norm*, however, is not the end of the story; more specific norms follow from it. In particular, in the contextualist framework, assertions are proposals to update the common ground of the conversation, which is viewed as an enterprise whose goal is generally to find out truths about the world. The common ground should thus generally track something like common knowledge, or at least common belief. It follows that the *Speech Act Norm* entails that an assertion is generally permissible only if the assertion

---

\(^{23}\)As in e.g. Williamson 2000; Lackey 2007. It doesn't matter for present purposes which of these mental states is the relevant one.

\(^{24}\)One reaction to these observations is to argue that performative ‘assertions’ are not really ‘assertions’ in the fullest sense: we should reserve ‘assertion’ for the class of speech-acts which are governed by doxastic norms. This is a terminological dispute. I have no quarrel with this alternate reaction to these data; the important point is that, either way, we will make sense of how speakers assert epistemic modal claims by assimilating them to performatives.
contributes to coordination on common knowledge or belief; that is, only if, should that assertion be commonly accepted, the common ground will track the knowledge of the interlocutors at least as well as if it were rejected. Call this corollary of the *Speech Act Norm* the *Common Ground Norm*.

The *Common Ground Norm*, in turn, entails a limited norm of the kind familiar from the literature on norms of assertion: the *Common Ground Norm* says that, for any sentence whose content is exclusively about a state of affairs which does not depend on the assertion of the sentence, one should generally assert it only if the speaker already knows or believes its content (with the exception, perhaps, of particular cases in which the common ground is not aiming to track common knowledge or common belief). Non-performative assertions will thus be governed by a familiar norm, which says that you generally have to know or at least believe the content you assert.

But when it comes to performative sentences—sentences whose content is about a state of affairs which depends on the assertion of that sentence—nothing in the *Common Ground Norm* entails that one must know or even believe the content of the assertion when asserting it. All that the *Common Ground Norm* says is that one's claim should generally contribute, or at least not detract, from the good epistemic standing of the conversation. It follows that performatives should generally be such that, *if they are accepted*, they will be commonly known, a mild and broadly plausible constraint. But it doesn’t follow that performatives have to be known—or believed—antecedently: one can assert risky performatives, performatives the speaker doesn’t know will be accepted (and thus doesn’t know will be made true), provided their aim is a permissible one.

This account of the norms of conversation makes sense intuitions both about when ordinary, non-performative claims can be asserted, as well as when broadly performative claims can be asserted. And, if epistemic modal claims are performatives, as prospective contextualism maintains,

---

25 Among other things, this suffices to explain the infelicity of performatives which command the addressee to do something that isn’t in their power, like ‘You will win the lottery’, said as a command to buy a lottery ticket. It also goes some way towards explaining the infelicity of Moore sentences containing performatives as a conjunct: ‘You’ll clean the rabbit cage but I don’t know you will’ cannot amount to common knowledge, even if it is accepted, provided the time interval relative to which ‘know’ is evaluated includes the prospective time. In theory, this doesn’t rule out performative Moore sentences if care is taken to ensure that the interval relative to which ‘knows’ is evaluated does not include the prospective time. Sentences like this seem to be ruled out on independent grounds, however; you shouldn’t tell someone to do something, and then say you aren’t sure if they will. To see this, note the oddity of ‘You have to clean the rabbit cage, but I don’t know if you will’, or ‘Clean the rabbit cage! I don’t know if you will’. It seems as though, in trying to get someone to do something, you must act as if you are confident that they will. Why this is true is another question, one whose answer is not obvious to me.
then, given the normative regime that governs performatives in general, we can resolve the normative concern about prospective contextualism. According to prospective contextualism, one can often assert modal claims without knowing their content to be true. But the fact that, according to prospective contextualism, modal claims are performatives, allows us to give a disciplined account of why this should be so: they are in the first instance used to do things—make proposals about what entailment and compatibility properties the common ground should have—and thus that they are governed by norms which look, in the first instance, at those actions, rather than at agents’ relation to the truth conditional content of those claims.

This also helps us say a bit more about what prospective contextualism predicts regarding when a modal claim can be made. According to the Common Ground Norm, a principle normative concern regarding assertions is whether they contribute to the epistemic coordination of the agents in conversation. Thus an assertion of “Must $p$”—a proposal to make $p$ common ground—will generally only be acceptable if one knows (or at least believes) $p$. This seems broadly plausible (though see the next chapter for further discussion). And an assertion of “Might $p$”—a proposal to ensure that the common ground is compatible with $p$—will, in general, be acceptable only if it contributes to the conversants’ coordination on their knowledge. It is somewhat harder to say just what this amounts to, but it will be something like: the speaker must know that $p$ is a possibility which should be taken seriously in the conversation—perhaps because she has some evidence in favor of $p$, or in any case knows that her evidence is compatible with $p$, and believes that her interlocutors should know this too. Again, these constraints make sense of our intuitions about when a modal claim can be made, and (therefore) what considerations we take into account when deciding whether to make modal claims.\(^{26}\)

The performative structure prospective contextualism attributes to epistemic modal claims thus furnishes a principled response to the normative challenge outlined above.

It may be objected at this point that, although the prospective contextualist framework is technically within the bounds of contextualism, it has abandoned one of the principle attractions of

\(^{26}\)Apparent exceptions to these generalizations come from ‘exocentric’ uses of modals. I will not discuss those uses at length here, but I assume that those uses are somewhat exceptional, involving either free indirect discourse of some kind, or a tacit modal operator, either of which would suffice to explain our use of them.
contextualism: namely, that in a contextualist framework, we can view assertions as coordination on the speakers' knowledge and beliefs; if you can assert propositions without believing them, the thought goes, we lose this attractive feature of contextualism. But this is mistaken. The normative framework I have sketched here still views conversation as (among other things) aiming at the coordination of speakers' knowledge and beliefs. It just turns out that—when the truth of a given content depends on the speech act of asserting that content—you need not know or believe that content in advance of it being accepted, in order for your assertion to contribute towards coordination of the speakers' information.

2.5 Comparisons

This discussion shows that prospective contextualism's treatment of epistemic modal claims is more plausible than it might first seem: provided we pay careful attention to the phenomena of performativity in natural language in general, we can make sense of the way that speakers assert, and negotiate about, epistemic modal claims within the prospective contextualist framework. In this section, I will briefly compare prospective contextualism with three competitor approaches. The aim, again, is not to present a decisive argument in favor of prospective contextualism, but rather to bring out some considerations that might motivate capturing the guiding observation within the contextualist framework, rather than within one of the alternative frameworks which has been proposed in recent years—considerations which I believe show prospective contextualism to be well worth further exploration.

I will discuss three approaches, each of which aims to capture the guiding observation by rejecting some part of the contextualist framework. In particular, each rejects—in different ways—the claim that a modal claim's content at a context is a proposition, and that the assertion of a modal claim is a proposal to add its content to the common ground. I will highlight two ways in which I believe this leads to a less satisfying theory of the dynamics of epistemic modality than the one prospective contextualism gives us. First, the resulting theories lose contextualism's hygienic division of labor between semantics and pragmatics. Second, and more worryingly, the resulting
theories, insofar as they do not treat the evolution of conversation as the incrementing of information to the common ground, no longer provide a fully explicit model of the dynamics of modal assertions; and rendering those theories fully predictive would require further ad hoc amendments. Nothing I say here is meant to be a decisive objection to any of these theories; I rather aim to highlight the benefits of staying within the contextualist framework, and the theoretical debts that are incurred when we leave it.

The first approach I will discuss is one way of fleshing out a proposal sketched in Stalnaker 2014. Stalnaker's proposal is similar to mine (and mine is indebted to it in obvious ways). Rather than building prospectivity into the meaning of epistemic modals, however, Stalnaker proposes that epistemic modals have as their meaning a simple quantificational structure, which is then coupled with a special force rule. This force rule specifies that an epistemic modal claim is to be interpreted as a proposal to make the prospective context set (the set of worlds compatible with the common ground) verify the modal claim: i.e. as a proposal to render the prospective context set such that, if the modal claim is interpreted with that set as the modal's domain of quantification, the claim comes out true.

The second approach is the expressivist approach, particularly as developed in Yalcin 2007, 2011, 2012a. On this view, an assertion of a sentence is a proposal to make the context set (the set of worlds compatible with the common ground) accept that sentence, where, for atomic sentences, a set of worlds accepts a sentence just in case either the sentence is non-modal and the set of worlds entails its content; or the sentence is modal and the modal claim is true when the set of worlds is taken as its domain of quantification.

The third is the update semantics developed in Veltman 1996. This view is given in a dynamic framework in which sentences denote functions from contexts to contexts, rather than propositions (where a context here is interpreted just as a set of possible worlds, in particular as the context set).

"Might p" denotes the function which takes an input context and either leaves it unchanged, if p
is compatible with that context; or else returns the empty set (crashes). "Must p" will thus be a function which takes an input context and either leaves it unchanged, if p is entailed by that context; or else crashes to the empty set.

All three of these approaches depart from the contextualist framework in obvious ways. Proponents of these approaches will, of course, maintain that that is as it should be: conversation is just more complicated than the contextualist framework countenances. But I believe this divergence leads to important drawbacks. I will highlight two. The first is the somewhat abstract point that, insofar as they depart from this model, these proposals lose the neat story that contextualism tells about the interface between semantics and pragmatics. The second objection I will sketch is more concrete and, I think, more serious.

Consider first Stalnaker's account. This account nicely conforms to the guiding observation. But on this approach, an epistemic modal claim is associated with a distinctive force rule. This multiplication of force rules adds complexity to the pragmatic framework which simply looks unmerited, given that we can do without it, as in the prospective contextualist framework. More seriously, it entangles our pragmatic theory with lexical semantics in an unattractive way: on Stalnaker's view, what triggers the distinctive prospective force rule is a claim's being an epistemic modal claim. This pragmatic rule thus makes reference to specific lexical items—modal words—rather than to the output of the semantic operations.\footnote{See Szabó 2016 for related criticism.} In other words, the pragmatics has to peer into the compositional semantic machinery before knowing what to do with a given content.\footnote{Importantly, this is in part because the meaning of modal claims will be structurally just like other propositions; compare the situation with questions or imperatives, which many believe are associated with distinctive force rules, but also have contents with distinctive semantic types.} This blurs the division between pragmatics and semantics: the former is meant to be a theory of what we do with language, independent from specific conventions of language use. Perhaps a proponent of a view like this will simply maintain that this modular view of pragmatics is wrongheaded. But I think that given the option—an option which prospective contextualism makes available—we are better off maintaining a careful division between the theory of the conventions of language use, on the one hand, and the theory of what we do with contents, on the other. A closely related point concerning Stalnaker's account concerns sentences which involve both modal and non-modal constituents. Consider a con-
junction with the form \( p \text{ and might } q \). What update rule should apply to these sentences? The prospective update rule simply does not make sense in application to the first conjunct; but the ordinary update rule will bleach out the proposal-like quality of the second conjunct. It looks like to make sense of assertions like this, we need a single update rule which applies to both modal and non-modal sentences. Similar points extend to disjoined modals, and so on. In order to make sense of these complex sentences containing epistemic modals, we need a single force rule for both modal and non-modal updates.

This is not to say that there is not something right at a descriptive level about Stalnaker’s characterization of the force of modal claims. A comparison with performatives like ‘This afternoon, you’ll be cleaning the rabbit cage’ is, again, helpful. There are prescriptive uses of this sentence, and also descriptive uses. These uses have different effects, and are appropriate in different cases; and, at the level of descriptive taxonomy, it is helpful to distinguish them. But we do not want to say that these uses are associated with distinctive force rules from the perspective of formal pragmatics (or semantics). Instead, we want a unified force rule which predicts that which of these different uses (prescriptive and descriptive) is brought out depends on varying background conditions (the normative authority of the speaker, their relationship to the addressee, and so on). Things are parallel for epistemic modals. While there is, at a descriptive level, something right in saying that modal claims have a distinctive force, in that they have a characteristic update effect of making a proposal about how the prospective common ground will look, we should not encode this in our pragmatic system; if we are to have any hope of making sense of even simple embeddings of epistemic modals in a principled way, then, from the perspective of our general pragmatic system, we need a single unified force rule for both modal and non-modal claims. Together with an appropriate account of the assertoric content of modal claims, the relevant descriptive generalization of the force of modal claims will then fall out as a consequence, rather than being stipulated in the system from the start.

There is a sense in which expressivism avoids this criticism by conforming to the guiding observation with just one update rule: make the context set accept the assertion in question. This rule does not mention lexical items. But the underlying update mechanism is still disjunctive: non-modal claims add their content to the common ground, while some other kind of adjustment is required for
modal claims. The resulting model is ad hoc. Why would interlocutors care about acceptance, in this disjunctive sense? How should we think about conversation, if it is all about acceptance? It is not clear that a theoretically satisfying account of this system can be given.

Consider, finally, update semantics. Update semantics, on the face of it, does not answer to our guiding observation about what ‘might’-claims are used to do. In the update semantics, ‘Must p’ can only be used to describe the context as it already stands, not to change it. Likewise, ‘Might p’ can only be used to claim that p is already entailed by the common ground, not inform one’s interlocutors that p is true. To make these problems dramatic, suppose that you think that John is in the pub. That means the context is compatible with the proposition that John is in the pub. Suppose I know John is in his office. The update semantics seems to predict that I should be able to assert ‘John might be in the pub, but he’s not.’ The update semantics predicts that this should sound roughly equivalent to ‘Although you think Bob is in the pub, he’s not’: I am simply noting, first, that the context leaves open that John is in the pub; and, second, that he is not in the pub. But the first sentence doesn’t sound at all like the second; as Wittgenstein (1953, II.x.110) and many since have observed, sentences like the first sound bad, whereas the second is perfectly fine. This fact can be seen as just another way of motivating our guiding observation. If ‘Might p’ is a proposal to make p compatible with the common ground, then of course (on any plausible semantics of conjunction) it will be incoherent to conjoin it with ‘Not p’, a proposal to ensure that p is not compatible with the common ground. By contrast, if ‘Might p’ is just a description of the common ground as it already stands, as on the update semantics, then there will be no incoherence here.

Proponents of update semantics have moves available in response to rule out assertions like this. Those moves, however, result, once more, in theoretically unsatisfying, ad hoc pragmatic systems. Two natural responses hold that you should only assert p if your belief state ‘supports’ p, where a state supports p just in case updating it with p leaves it unchanged; or that an assertion is a proposal to make the context support what is asserted (see Willer 2013). These moves provide technical solutions to our problem, since no information state will support ‘Might p and not p’—and more broadly speaking, a way to bring the update semantics in line with the guiding observation. But

---

33 This is brought out clearly in Rothschild (2011)’s presentation of expressivism.
these moves do not provide a very satisfying theoretical solution. After all, why should you only assert things that are supported by your belief state? Why should an assertion be a proposal to make the context support what was asserted, as opposed to simply a proposal to update the context with the asserted content? Given its semantics alone, 'Might p and not p' seems like it should be a perfectly good, and sometimes useful, thing to say. We can enrich our pragmatic system to get the results we want in the update semantics, but the enrichments are ad hoc, and, once again, custom made to interface in the right way with the semantics for epistemic modals.34

All three of these views, insofar as they depart from contextualism in order to capture the guiding observation, thus must complicate their pragmatic theories in ad hoc ways. This helps bring out what is attractive about making sense of the behavior of modals within the austere contextualist theoretical framework.

A second and somewhat more pressing problem for these divergences from contextualism is that they lose much of the predictive power of contextualism. I'll focus on expressivism in bringing this point out, but these points extend, mutatis mutandis, to the other two views. The issue is that, insofar as they no longer model conversation as the incrementing of information, these views end up telling us very little about how conversation actually evolves. On the expressivist view, an assertion of p is a proposal to make the context set accept p. But there are many ways for a context set to accept p: if p is non-modal, then any set of worlds where p is true will accept it; if p is modal, say of the form 'Might q', then any set of worlds which includes a q-world will accept it. Expressivism does not tell us anything about which set the updated context will be; in other words, it does not tell us anything about the relation between the antecedent context set and the updated context set. What we want, intuitively, is for the posterior context set to keep as much information from the antecedent context set as possible, while also coming to accept the new assertion. It may be thought that expressivism can capture this effect simply by appealing to general principles about belief revision. But this is wrong: since updating with an assertion, in the expressivist framework, does not in general amount to learning a piece of information—since sentences do not express propositions,

34Beaver (1992); Willer (2013) give variations on the dynamic framework which help with this problem, but they are still vulnerable to a slightly more complex form of it. For instance, Willer's approach predicts that once the common ground is updated with 'Might p', any subsequent update with 'Must not p' will always induce a crash to the empty set. To avoid this, once more, we would need a special pragmatic rule.
relative to contexts—expressivism, unlike contextualism, cannot appeal to general principles about belief revision to constrain the evolution of contexts.

To make this a bit more concrete, suppose the antecedent context set is $c$, and that $c$ is compatible with $p$ and $\bar{p}$. Suppose now that \textquoteleft\textquoteleft Must $p$\textquoteright\textquoteright\ is asserted. The expressivist framework tells us that the updated context set will accept \textquoteleft\textquoteleft Must $p$\textquoteright\textquoteright. This just amounts to the updated context set being a subset of $p$. But expressivism says nothing about the relation between $c$ and this posterior context set. What we want, intuitively, is for the posterior context set to be the largest subset of $c \cap p$ which satisfies the logic of common ground. But how will the expressivist achieve this? Since what is being 'added' to the context is a modal claim, not an ordinary factual assertion, no general considerations about belief revision will get us to this result.

A natural first thought in response to this problem would be to model a context as a set of information states, rather than a single information state, and then model update simply as intersection: the posterior context is just the set of all information states from the antecedent context which accept the asserted sentence. But this is a non-starter. For instance, suppose that \textquoteleft\textquoteleft Might $p$ and might not $p$\textquoteright\textquoteright\ is asserted; then the context will contain only information states compatible with both $p$ and $\bar{p}$. Now suppose sometime later that $p$ (or \textquoteleft\textquoteleft Must $p$\textquoteright\textquoteright) is asserted. There will be no information states in the context which accept this assertion, and so we will have a 'crash' to the empty context. Once more, we would have to superadd special pragmatic rules to capture even basic conversational dynamics like this. We run into a similar problem if we model the context set as a set of pairs of worlds and information states, and try to model update as set intersection.

The problems I am pointing to are not at heart technical problems. It is technically possible to augment the expressivist system with further update rules which will make the correct predictions. But I do not see a way to motivate such rules on general, theoretically satisfying grounds, rather than as \textit{ad hoc} stipulations. A way to dramatize this: if \textquoteleft\textquoteleft Must $p$\textquoteright\textquoteright is associated with the expressivist semantics and pragmatics, we could easily imagine that assertions of \textquoteleft\textquoteleft Must $p$\textquoteright\textquoteright would be used to \textit{reset} the context set, so that the context set comes to be just the set of all and only worlds where $p$ is true. Such a speech act could be quite useful. And we could augment the existing expressivist system to predict just this. Now, this is not in fact what we observe; but nothing in the expressivist
system rules out this update rule.

Put another way, the worry is as follows. On Yalcin’s account, the semantic value of a non-modal sentence \( p \) is the set of information-state/world-pairs such that the world makes \( p \) true. By contrast, the intension of \( \Box p \) is the set of information-state/world-pairs such that the information state entails \( p \). From a formal point of view, these two objects are entirely different. And yet we want updates with \( p \) and with \( \Box p \) to have essentially the same effect. But why should updates with these two different kinds of formal object have the same effect? This is something that we could, from a technical point of view, stipulate in the system (and likewise for the whole array of possible updates)—though this hasn’t been done, and it remains to be seen how exactly this would work. But, in any case, it would be just that: a stipulation. By contrast, if modal claims just express propositions—i.e. communicate information—as on the view I am advocating, then, to make sense of the update effects of modal claims, we need to advert to nothing more than an independently motivated theory of belief revision (in particular, this will predict without further stipulation that \( \Box p \) and \( p \) have essentially the same update effect; see again Appendix B for a more careful demonstration of this).

Again, similar worries extend for the update semantics, as well as Stalnaker’s semantics. By contrast to these theories, prospective contextualism, again, gives a fully deterministic account of the dynamics of conversation, at least once we have coupled it with a general, independently motivated theory of belief revision (see again Appendix B). In the prospective contextualist framework, modal claims, like all assertions, express propositions. So updating with a modal claim just is updating with an ordinary proposition—in other words, coming to accept a piece of information. We can thus advert to general considerations about belief revision to give a fully deterministic, and independently motivated, theory of how such updates will go.

Departing from contextualism in order to capture the guiding observation thus incurs serious theoretical debts, debts which prospective contextualism avoids. This does not yet yield a decisive objection; these costs may end up looking acceptable in light of other benefits these theories bring.\(^{35}\)

\(^{35}\)I do not think this has been shown to be true, though to show this would require discussion beyond the scope of this chapter. For instance, issues about embedding which have been taken to motivate the update and expressivist approaches can, I think, be better handled in the contextualist framework, and those approaches face new, serious embedding issues.
In any case, however, prospective contextualism shows clearly that we cannot justify those costs on the basis of the fundamental dynamics of epistemic modality alone. Prospective contextualism shows that it is possible to capture the guiding observation within the contextualist framework, treating assertions of modal and non-modal claims alike from the point of view of semantics and pragmatics, and thus preserving the simple contextualist vantage point on conversation, with its fully explicit and well-motivated theory of the dynamics of conversation.

2.6 Implementation

Prospective contextualism, again, is a theory about the assertoric content of modal claims: what proposition they contribute to the common ground when asserted and accepted. Prospective contextualism as I have presented it remains agnostic about the compositional semantics of those claims: i.e. about how modal claims end up with the assertoric content they have, as a result of compositional semantic machinery interacting with pragmatic considerations.36 This is not my main interest here; there are a number of different semantic routes to prospective contextualism, and the choice between them will depend on considerations involving embedded modals which are beyond our scope here. Before concluding, though, let me very briefly sketch what I take to be the most promising route.

First, let me note that prospective contextualism is broadly compatible with the standard semantics for modality sketched at the outset, on which modal claims quantify over a set of ‘accessible’ worlds. The main innovation in my view is about what set of worlds counts as accessible. However we capture this fact, we should not need to depart much from the standard semantics for modals.

The most obvious approach would be to hardwire the prospectivity into the content of modal claims, by stipulating in the semantics that a context c determines an accessibility relation which takes every world to the set of worlds compatible with the prospective common ground, at that world, of the actual conversation’s counterpart there. This approach would serve our purposes, but it would go wrong when it comes to embedded modals. This approach would maintain that "Might p" and "The prospective common ground is compatible with p" have the same semantic content. But

---

36 On the distinction between assertoric and semantic content see e.g. Ninan 2010.
this is plainly false, as can be seen by embeddings of the kind discussed in Yalcin 2007: sentences with the form "Suppose p and might not p" are felt to be infelicitous, but not so for sentences of the form "Suppose p and the prospective common ground is compatible with p". This means that, even though we are maintaining that "Might p" and "The prospective common ground is compatible with p" have the same assertoric content, we cannot identify their compositional semantic contents; we need a more sophisticated approach.

A second natural approach would be to take the standard modal semantics as our compositional semantics, and then assume that the prospective interpretation happens entirely at the pragmatic, 'post-semantic' level. This does not seem implausible to me, but it again leaves unexplained embedding data, including Yalcin's data, as well as data of the kind discussed in Groenendijk et al. 1996; Aloni 2001.

A more plausible approach, I think, walks a line between these two. The idea is to build the prospective contextualist interpretation on top of the semantic framework sketched in the first chapter. In that framework, since the local context for unembedded claims is just the set of worlds compatible with the common ground, unembedded epistemic modals will be presupposed to quantify over a subset of the common ground. Since the worlds compatible with the prospective common ground will generally be presupposed to be a subset of the worlds compatible with the present common ground (since it is generally presupposed that conversants will gain, not lose, information), they will constitute an admissible set of accessible worlds. General pragmatic considerations would be responsible for ensuring that, for unembedded modals, this is the particular subset selected. Absent other clues, this will be an obvious default, since it provides a strong and natural interpretation of what speakers might be trying to do with modal words. The benefit of this approach over hardwiring prospectivity into the semantics is that it gives us a better framework for making sense of embedding data; as we saw above, the local context can be shifted by embedding operators, explaining a variety of data involving embedded epistemic modals. In particular, this approach predicts that, while an unembedded assertion of "Might p" will generally be interpreted as having the same content as an unembedded assertion of "The prospective common ground is compatible with p", these two expressions will *embed* differently, since they have different compositional semantic values.
This approach, further, predicts that prospective contextualism will be a default interpretation, but that there may be situations in which other bodies of evidence are sufficiently salient that we interpret modal claims differently. This seems like a plausible way of walking the line between, on the one hand, the guiding observation that modal claims are generally used to negotiate about what possibilities to treat as live; and, on the other, the observation that, provided suitable contextual set-up, they may be used simply to describe someone’s state of mind, as in the ‘stand your ground’ cases discussed in §2.4.1, or in ‘exocentric’ readings of modals.\(^{37}\)

Although this is not the place to discuss embedded modals, let me note that prospectivity will still play an important role in their interpretation. Thus e.g. "If p, then must q" will, as a default, be interpreted as saying that q is true in all the p-worlds of the prospective context. Just as with unembedded modals, this allows us to interpret modals embedded under connectives in a way which makes sense of what speakers are doing with them (in this case, negotiating about whether to make all the p-worlds in the context q-worlds); thus in particular an assertion of "If p, then must q" will have the same update effect as an assertion of "Not p or q" (despite having a different semantic value).

Spelling out these ideas in detail goes well beyond the scope of this chapter, whose goal is to show what assertoric content we must assign to modal claims in order to conform to the guiding observation; but this should suffice to give a sense of how we might implement prospective contextualism in a semantically respectable way.

### 2.7 Conclusion

Prospective contextualism makes sense of the fundamental dynamics of epistemic modality within the bounds of contextualism. This shows that, pace much recent work, the dynamics of epistemic modality do not present an insoluble challenge to the contextualist framework: we can preserve contextualism’s elegant model of communication as information transfer, and its corresponding division of labor between semantics and pragmatics, while making sense of the guiding observation about how speakers use modal claims to negotiate about what possibilities to treat as live.

\(^{37}\)See also e.g. Egan et al. 2005; von Fintel and Gillies 2008; Kratzer 2012a, for related cases like this.
That we can do so does not mean that we should; whether prospective contextualism is the correct model of epistemic modality is another matter, one I have not taken a stand on here. I have addressed an obvious objection to prospective contextualism, arguing that the performative structure which prospective contextualism attributes to modal claims is widely attested in natural language, a fact which helps us make sense of the norms of modal assertions within the prospective contextualist framework. I have also brought out some theoretical debts which competitor approaches must pay insofar as they depart from contextualism. I have not tried to show decisively that prospective contextualism is the correct model of epistemic modality. My goal, instead, has been to show that contextualism cannot be rejected on the grounds of its putative inability to capture the guiding observation, and that the resulting version of contextualism—prospective contextualism—is well worth serious further exploration as a theory of epistemic modality.

To see whether prospective contextualism is the correct theory of epistemic modal claims, we would have to explore a broad range of further questions. Let me highlight two particularly important further questions involving epistemic modals, and what resources prospective contextualism might have to address them. The first comes from embedded epistemic modals. I believe that prospective contextualism can be implemented in a way which accounts for the embedding behavior of epistemic modals, by building on the semantic framework developed in the first chapter, and therefore I do not think that embedding data pose any challenge for prospective contextualism.

The second issue concerns certain kinds of reactions to modal claims which have motivated relativism about epistemic modals. If I overhear Sue saying, ‘John might be in his office’, and I know that John isn’t in his office, then relativists claim that it is appropriate to respond to Sue by saying ‘No’ or ‘That’s not true’, even if I am not part of Sue’s conversation. This is a prima facie puzzle for prospective contextualism, which predicts that Sue’s assertion is a proposal to leave it compatible with her common ground that John is in his office, not with my common ground. Puzzles like this have led some to reject contextualism, arguing that truth is determined relative not only to a world but also to a judge or information state supplied by the context of assessment.\(^{38}\)

This challenge is complicated, both theoretically and empirically. I will not attempt to fully

\(^{38}\)See e.g. Egan et al. 2005; Stephenson 2007a,b; Lasersohn 2009; 2.
address it here, but let me make two brief remarks about it. The first is that there are theoretical reasons to resist a move to a relativist framework, reasons similar to those I gave above for resisting the move to dynamic and expressivist frameworks. When we move to a relativist framework, we must abandon the view of conversation as figuring out, together, which world we are in, since, from a relativist perspective, different conversants are in different “worlds” (different relativist points of evaluation). This makes it hard to see what the point of conversation is and how we should model its dynamics; although there have been some attempts to answer these challenges, I do not think they have been successful, though there is not space here for an adequate discussion.\(^{39}\) Again, it may be possible to answer these objections in a satisfying way, but these issues at least make it well worth our while to explore whether the move to relativism can be resisted within a contextualist framework.

The second point is that, by giving a successful contextualist model of the basic intra-contextual dynamics of modal claims, prospective contextualism may provide a theoretical foundation for a successful model of the dynamics of the cases which have motivated relativism. Whether it can do so depends on what exactly the empirical picture is, which turns out to be much more complicated than relativists first suggested. Recent empirical work in Knobe and Yalcin 2014; Khoo 2015 has shown that, contrary to the main claim in the relativist literature, subjects are not actually much inclined towards cross-contextual truth-value contestations: that is, if I overhear Sue saying that John might be in his office, subjects do not find it to be particularly appropriate for me to reply ‘That’s not true’ (they find this to be a markedly less appropriate response than in corresponding

\(^{39}\)Cf. \textsuperscript{?} for related criticism. See Egan 2007; Stephenson 2007a for attempts to answer these challenges, neither of which I find satisfying. Egan proposes that “Might \textit{p}” is evaluated relative to world-individual pairs, and true just in case \textit{p} is compatible with what’s in the ‘epistemic reach’ of the individual at that world. He models the common ground as a set of world-individual pairs, and models assertions as adding information to the common ground by intersection. For this to make sense, Egan shows that there must be a presupposition in place that all conversants are relevantly the same with respect to their epistemic reach (otherwise, assertions will end up ‘stranding’ conversants). Egan argues that this is plausible, since it is just part of the notion of epistemic reach that if I’m talking to someone for whom \textit{p} is entailed by what’s in their epistemic reach, then \textit{p} is entailed by what’s in \textit{my} epistemic reach. But such a strong notion of epistemic reach leads to trouble. Egan argues that you must believe what you assert; thus to assert “Might \textit{p}” in Egan’s system, a speaker would have to believe that \textit{p} is compatible with what’s in her epistemic reach; it follows from this strong notion of epistemic reach that she would have to believe \textit{p} is compatible with what \textit{everyone} in the group knows (since whatever someone knows is, presumably, entailed by their epistemic reach). But, like the group contextualism considered and rejected in §2.2.2, this makes the assertion of ‘might’-claims implausibly demanding. Stephenson’s account avoids this issue by arguing that the content one must believe in order to assert “Might \textit{p}” is substantially weaker than the content that gets added to the common ground. But this move is, as far as I can tell, totally \textit{ad hoc}, and thus theoretically unsatisfying.
non-modal cases). That is exactly in line with the predictions of prospective contextualism, and contextualist approaches to epistemic modals more broadly speaking. But that same work shows that expressions of disagreement ('No, Joe is at home') and subsequent retraction ('Scratch that, Joe is at home') are generally felt to be reasonably acceptable in contexts like the one described.

Khoo (2015), however, has provided a persuasive account of how a contextualist account can make sense of these facts. The basic idea is that, if a contextualist account can predict (what we have called) the guiding observation, then 'might'-claims will be felt to be proposals to make their prejacents compatible with the common ground, and cross-contextual or retrospective disagreement can target exactly this proposal. To reject or retract a 'might'-claim is, on this approach, to signal that one does not stand behind the proposal made by the 'might'-claim in the first place. Thus a response of the form ‘No, Joe is at home!’ can be interpreted as, essentially, saying that it is a bad idea to leave open the possibility that Joe is in his office—what Sue is proposing in saying ‘Joe might be in his office’. Khoo provides a careful exposition of this idea; I refer readers to the paper for details. Crucially, Khoo notes that this kind of response is contingent on finding a version of contextualism which predicts the guiding observation, observing that ‘No contextualist theory has attempted to predict the dynamic update effects of uttering epistemic modal sentences’ (Khoo, 2015, p. 529). But prospective contextualism does just this. Thus, in concert with a story about (dis)agreement along the lines Khoo gives, prospective contextualism provides a promising platform for responding to the relativist challenge to contextualism.

There is, of course, much more to explore with respect to this complicated empirical domain, but this discussion shows that, insofar as it accounts for the guiding observation, prospective contextualism provides substantial new resources to the defender of a contextualist theory of conversation.

40To the degree that subjects still find it to be slightly appropriate, there remains something to be explained here. Two possibilities, both of which seem plausible to me: first, expressions like 'That's (not) true' can sometimes serve simply to register broad (dis)agreement, rather than to contest a truth-value. (Evidence for this comes from the fact that expressions like this are used reasonably frequently in response to questions. 'Have you done a follow-up examining cross-cultural variations in this result?' ‘That’s true, that’s an excellent line to pursue.’) Second, pronominal expressions like 'That' or 'What S said' may in some cases refer to something other than the proposition expressed by S. It is well known at this point that pronouns in general have “sloppy” uses (see Karttunen 1969), and it seems perfectly plausible that this goes for these pronominal expressions, too: they may be able to pick out something like a function from assignments of contextual parameters to propositions, rather than a proposition. There are obvious limits to the extent to which this is possible, limits which I think are fairly straightforward to make sense of within standard theories of the $\varphi$-features of pronouns, but that is a topic for another time.
against relativist challenges.

In concluding, I would like to suggest that the framework I have sketched here can be generalized to analyze not just epistemic modals but also a wide variety of other constructions that we use to negotiate matters that depend, in part or in whole, on parameters of the context. The extension to probability modals like 'Probably', which have much in common with epistemic modals, will be straightforward: these can be seen as making claims about what the contextually provided measure structure will be like at the prospective time; provided the identity of that measure structure depends on what it is commonly accepted to be, then probability modals can be used to negotiate about what measure structures to coordinate on. Deontic modals can receive a similar treatment. The standard theory of deontic modals faces a similar puzzle to that raised at the outset for the standard theory of epistemic modals: how do speakers use deontic modals to negotiate about what to do? If I say we should go to Chinese, and you disagree, we are clearly not disagreeing about what norms are accepted by me, nor about what norms are accepted by everyone. Instead, it looks like we are disagreeing about what norms to accept. We can capture this by saying that we are making assertions about what norms will be accepted at the prospective time; provided the identity of those norms depends on what is commonly accepted about them, deontic modals can be used to negotiate about what norms to coordinate on.

A similar treatment may be available for a variety of other phenomena, such as negotiation about standards of vagueness, matters of taste, and performatives of the kind discussed in Austin 1979, like 'I promise'. It is tempting to think that phenomena involving negotiation of the sort these involve take us beyond the contextualist framework, since in these cases we seem not to be describing how things are vis-à-vis some contextual parameter, but rather proposing how they ought to be. But the contextualist framework can make sense of this if we take these constructions to describe the way that parameter will be. Provided that the identity of the parameter depends in the right way on what the interlocutors accept it to be, constructions which describe how the parameter will come to be will amount to performative proposals about how the parameter ought to be.
Chapter 3

What ‘must’ adds

This chapter explores the role of ‘must’ in conversation. Assertions of ‘Must p’ and assertions of p alone seem to have the same basic goal: namely, coming to agreement that p is true. Yet their felicity conditions differ in subtle ways. I argue for a new characterization of those differences, and a broadly pragmatic account of them. My central claim is that an assertion of ‘Must p’ is a proposal to accept p, but to do so based on a shared argument, rather than based on the speaker’s epistemic authority. I show that the requirement that the speaker’s evidence for p be indirect follows pragmatically from this constraint. I argue that this approach accounts for the sense in which ‘must’ is strong (it has a strong update effect), as well as the sense in which ‘must’ is weak (the speaker is proposing an update, not on the basis of her own epistemic authority, but rather on the basis of a shared argument).

3.1 Introduction

Consider the following pair of sentences:

(1) It must be raining out.

(2) It is raining out.

1An excerpt of this chapter will appear as Mandelkern 2017.
Intuitively, an assertion of (1) and an assertion of (2) have the same basic aim: they are both proposals to accept that it is raining out. Once an assertion of (1) has been accepted, interlocutors are disposed to accept the content of (2): that it is raining out. Thus (1) seems to be as strong as (2). But it does not seem to be stronger than (2): it is very strange to assert (1) after (2) is already accepted, as witnessed by the oddness of (3):²

(3) ??It’s raining; and moreover, it must be raining.

This suggests that assertions of (1) and (2) carry the same basic information. Yet the conditions under which they can be felicitously asserted differ in subtle ways. Suppose that Jane is in a windowless room, and sees her colleagues come in with wet umbrellas. Then she can assert either (1) or (2). But now suppose that Jane is looking out a window at the rain. She can still assert (2), but an assertion of (1)—"It must be raining out"—would be decidedly odd.³

Generally speaking, there exists a systematic difference between the conditions in which one can felicitously assert a ‘must’-claim with complement p, versus the conditions in which one can felicitously assert p alone.⁴ The goal of this chapter is to account for this difference—both what it amounts to, and what explains it. This puzzle, which gets to the heart of a number of broad foundational questions involving epistemic modals and the dynamics of conversation, is known as Karttunen’s Problem.⁵

The main argument of the chapter comes in three parts. In §3.2, I get clear on the data: what the difference in felicity conditions between sentences like (1) and (2) amounts to. The main claim in the literature, which I call Indirectness, is that a ‘must’-claim is felicitous only if the speaker’s evidence for its prejacent is indirect, whereas its bare prejacent can be asserted whether the speaker’s

² ’??’ is used to mark felt infelicity, without suggesting anything about the source of the infelicity.
³ The example, and this way of setting up the puzzle, is from von Fintel and Gillies 2010.
⁴ A ‘must’-claim is a claim (an assertion, or that assertion’s content) containing an unembedded strong epistemic necessity modal: an expression such as ‘must,’ ‘it is necessary that,’ ‘it has to be that,’ ‘it can’t be that’, etc., read epistemically, unrestricted by attitude predicates, past tense, or other implicit or explicit restrictions. I will use ‘must’ as an exemplar of such modals, though let me emphasize that my claims here are not about the specific lexical item ‘must’, but rather about the class of words that have the meaning of epistemic ‘must’. I will use prejacent to refer to the clause or proposition that a modal takes as a complement; I assume throughout that ‘must’ takes a prejacent that is not itself a ‘must’-claim. I use ‘p’ as a sentence variable; ‘p’ denotes the proposition expressed by p (I suppress implicit relativization to contexts for readability). I move freely between talking about assertions of sentences and assertions of propositions.
⁵ Following von Fintel and Gillies (2010), who credit Karttunen (1972) with bringing the issue to attention.
evidence is direct or indirect. I argue that, while Indirectness is correct, there is another, equally important, generalization which plays a key role in solving Karttunen’s Problem: namely, that a ‘must’-claim is felicitous only if the speaker ensures there is a salient argument in support of the claim’s prejacent. I call this constraint Support. I provide a battery of cases to argue that Support is needed to characterize the difference in felicity conditions between ‘Must p’ and p, as well as experimental results which further confirm this hypothesis.

In §3.3, I show that once we have Support clearly in sight, we can derive Indirectness through general pragmatic reasoning. This reasoning reduces our judgments about the indirectness of ‘must’ to judgments about when a sequence of assertions is redundant. I argue that this approach is more explanatorily and empirically adequate than any other extant account of Indirectness.

In §3.4, I give an account of why Support arises in the first place. I argue against the few extant accounts, and then propose an account on which Support arises as a Gricean manner implicature, given certain assumptions about the way we interpret ‘must’-claims along the lines of those defended in Chapter 2.

In §3.5, I discuss how to account for the felt weakness of ‘must’. I argue that my account predicts that—although ‘Must p’ has a strong update effect—a speaker’s warrant for asserting ‘Must p’ may be weaker than her warrant for asserting p alone has to be (since ‘Must p’ is a proposal to accept p based on shared evidence, not the speaker’s epistemic authority), thus accounting for the intuition that ‘must’-claims are sometimes felt to be in some sense weaker than the corresponding non-modal claims. Finally, in §3.6, I discuss the question of how to extend my account to embeddings, before concluding in §3.7.

3.2 The data

I begin by getting clear on the difference in felicity conditions between assertions of ‘Must p’ and p.
3.2.1 *Indirectness*

The main claim in the literature is that this difference amounts to an *indirectness constraint*:6

*Indirectness*: A claim of "Must p" is felicitous only if the speaker’s evidence for p is indirect; a non-modal claim can be felicitous whether the speaker’s evidence for it is direct or indirect.

*Indirectness* is motivated with cases like (4) and (5):

(4)  [Watching the rain:]
   a. ??It must be raining.
   b. It’s raining.

(5)  [Seeing her colleagues enter with wet umbrellas:]
   a. It must be raining.
   b. It’s raining.

(4-a) is distinctly weird as compared with (4-b), (5-a), and (5-b). *Indirectness* is the most natural generalization to draw from data like these. It has been well-motivated in the literature, and so I will assume it is correct without further discussion.

More needs to be said about the questions of what counts as ‘indirect’ and what accounts for *Indirectness*, both questions which I return to below. For now, though, note that I do not assume that the concept of indirectness which plays a role in *Indirectness* neatly matches all intuitions about whether evidence is direct or not.7 For instance, reliable testimony is intuitively *indirect* evidence; but as von Fintel and Gillies (2010) observe, we must treat it as ‘direct’ when it comes to evaluating *Indirectness*, in order to predict the infelicity of assertions like (6-a) in most contexts:

(6)  [Tom tells Susie that it is raining. Susie says to Mark:]
   a. ??It must be raining.

---

7 Or categories which are encoded as grammatical markers of evidentiality in some languages; see e.g. Willett 1988, Aikhenvald 2004.
b. It’s raining.

Part of what follows will be an attempt to cash out and explain the relevant notion of indirectness.

### 3.2.2 Support

Does *Indirectness* exhaust the difference in felicity between ‘Must p’ and p? Most of the literature on Karttunen’s Problem has indeed focused exclusively on *Indirectness*. But a different thread in the literature has pointed to a further contrast in felicity conditions between ‘Must p’ and p: in making a ‘must’-claim, the speaker must ensure that an argument for its prejacent is salient to all the interlocutors.

**Support:** An assertion of a strong epistemic necessity claim, like ‘Must p’, is felt to be substantially degraded unless there is an argument for p salient to all the interlocutors. By contrast, a non-modal claim, or a claim with an epistemic modal auxiliary which is not a strong necessity modal is not felt to be degraded without an argument.

Support says, essentially, that a claim of the form ‘Must p’ needs a salient argument for p—it will be substantially degraded without such an argument—by contrast to non-modal claims, as well as epistemic modal claims of other strengths (like ‘Might p’, ‘Should p’, or ‘Probably p’). Part of Support was observed by Stone (1994), who called attention to the fact that ‘must’ requires an argument. This claim, however, has not received much discussion in the subsequent literature, and the contrast to modal claims of other strengths has not, as far as I know, been observed. The data that motivate Support are indeed less clearcut than those that motivate *Indirectness*. This is unsurprising: evaluating Support requires evaluating discourses as a whole, rather than single utterances, and it can be quite difficult to determine, in a given context, whether an argument has been made salient. Both these facts produce a fair amount of noise in evaluating Support. In the remainder of this section, I will argue that Support is indeed required to account for the difference in felicity conditions between ‘Must p’ and p, as well as the difference in felicity conditions between ‘Must p’ and modal claims of other strengths (the relevance of this latter observation will become clear later in the chapter, when we turn to the question of how to account for Support).

---

8It is taken up briefly in Murray 2014, Silk 2014, Swanson 2015, and Lassiter 2016.
Consider the following case:9

(7) Patch the rabbit sometimes gets into the cardboard box where her hay is stored. On his way out the door, Mark hears a snuffling from the box and thinks to himself, ‘Patch must be in the hay box.’ When he gets to school, Bernhard asks him how Patch is doing.
   a. [Mark:] ?? She’s great. She must have gotten into the box of hay this morning.
   b. [Bernhard:] Cute!

Suppose the conversation ends here, and assume that Bernhard doesn’t know anything about Patch’s set-up at Mark’s house, or in general anything which might help him figure out why Mark thinks that Patch was in the box of hay. There is something distinctly odd about this exchange. Intuitively, what Mark has said needs some more elaboration; either Mark should have proffered reasons to think that Patch was in the hay box, or Bernhard should have asked him for reasons, perhaps by saying, ‘Why do you say that?’ Here is a more felicitous version of (7):

(8) a. [As in (7), except Mark says] She’s great. I heard a snuffling from the box of hay on my way out—she must have gotten into the box.
   b. [Bernhard:] Cute!

Now suppose the conversation ends here. This exchange has none of the peculiarity of (7).

What does this show? First, note that a non-modal variant of (7) is perfectly fine:

(9) a. [As in (7), except Mark says:] She’s great. She got into the box of hay this morning.
   b. [Bernhard:] Cute!

The non-modal variant on (8) is likewise fine. The infelicity of (7) thus seems to be due to the use of ‘must’. But note that Mark’s evidence regarding Patch’s whereabouts is exactly the same in (7) as in (8). Moreover, Indirectness seems to be satisfied: Mark’s evidence about Patch’s whereabouts is certainly indirect, in whatever sense is generally relevant to evaluating ‘must’-claims. And yet

9I will include double question marks in this section, but note that the infelicity in question is attached to the discourse as a whole, and not to any one sentence.
(7) seems strange. So *Indirectness* cannot explain the contrast between them. By contrast, *Support* predicts precisely the pattern we observe: namely, that (7) will be relatively degraded—since no argument for the prejacent is made salient—while (8) and (9) will be substantially more acceptable.

Cases like this thus provide our first piece of evidence for *Support*. In a moment I will give a variety of cases which elicit similar intuitions. But first we should say more about what *Support* amounts to. First, what does an argument amount to? I will think of an argument for *p* in a particular context as a set of propositions which the speaker is commonly recognized to believe provides reason to believe *p*—either by deductively entailing its conclusion; by inductively supporting the conclusion; or by showing how the conclusion follows from what is already accepted.

Second, what does ‘salience’ amount to? I won’t say much about this, but a few features are worth noting. First, an argument need not itself be common ground, i.e. need not itself be commonly accepted in the conversation. One can felicitously assert an argument conjoined with a ‘must’-claim, even if the argument has not yet been (and never is) accepted by all the speakers (if Bernhard doesn’t believe me that I heard a snuffling from the box of hay, this does not render (8) infelicitous). The sense in which an argument *Γ* must be salient is rather that it must be common ground that the speaker takes *Γ* to provide reason to believe the prejacent of her ‘must’-claim, and that she is proposing to add *Γ* to the common ground. I will refer to an argument with this status as ‘salient’ or ‘shared’ or ‘publicly available’.

An important point about salience is that an argument can be salient without being made explicit, as in (10):

(10)  
[Bernhard and Mark are in the bunny’s room, and can both hear snuffling from the box of hay. Mark:] Patch must be in the hay box.

Here, the premise that merits Mark’s conclusion—that Mark can hear snuffling from the box—is salient, and the ‘must’-claim is acceptable.

Another noteworthy feature of the notion of salience in question is that the argument in question need not be salient *at the time of the assertion*; it can be provided shortly after the assertion (match-
ing patterns for other constructions that require something to be made salient, like the resolution of referents for demonstratives). Note, for instance, that (11) is felicitous:

(11)  
a. [As in (7), but Mark says] Patch must have gotten into the box of hay.
b. [Bernhard:] Why do you say that?
c. [Mark:] I heard her snuffling around when I was leaving.

3.2.3 More cases

We find further confirmation of Support when we turn our attention to a broader range of cases. Consider e.g. (12), adapted from Murray 2014:

(12) Sarah works in a windowless building. On her way to a meeting, she sees her coworker Jim enter the building, carrying a wet umbrella. Sarah concludes from this that it’s raining out. Sarah enters the meeting. Her colleague Thomas, who didn’t see Jim carrying a wet umbrella, asks, ‘What’s the weather like?’ Sarah responds:

a. ??It must be raining out.
b. It’s raining out.
c. It must be raining out; I just saw Jim come in with a soaking wet umbrella.
d. It’s raining out; I just saw Jim come in with a soaking wet umbrella.

Thomas replies: ‘Oh, too bad. Ok, let’s talk about the agenda for this meeting.’

As Support predicts, (12-a)—the variant with ‘must’, but without an argument—is odd, while the other variants (‘must’ with an argument and non-modal with or without an argument), are fine.

We find a similar pattern in (13):

(13) Jane is in her first year of college. She doesn’t have a clear sense of how she is doing in school. She meets with her professors, who tell her she is doing well; she thus concludes that she is doing okay. She goes in to meet with her adviser to talk about course registration. Her adviser doesn’t know about the conversations she’s had with her professors. Her
adviser asks: ‘So, how are you doing in your classes?’ Jane responds as follows:

a. I must be doing okay!

b. I’m doing okay!

c. I must be doing okay: I’ve spoken to all my professors and they told me I’m doing fine.

d. I’m doing okay: I’ve spoken to all my professors and they told me I’m doing fine.

[Her adviser replies:] ‘Good, I’m happy to hear that. Ok, on to our business for today: let’s discuss your registration for next term.’

Again, as Support predicts, (13-a) is marked, while the other variants are fine.

The next example illustrates the contrast of ‘must’-claims without an argument not only to ‘must’-claims with an argument and non-modal claims with or without an argument, but also to epistemic modal claims of other strengths:

(14) Two friends, Scott and Mark, are discussing summer plans. Scott asks Mark: ‘Do you think you’d be free to go fishing in a few months, say in the first week of September?’

Mark responds:

a. ??Yeah, I must be off work that Monday. Where would you want to go?

b. Yeah, that Monday is Labor Day, so I must be off work. Where would you want to go?

c. Yeah, I’m off work that Monday. Where would you want to go?

d. Yeah, that Monday is Labor Day, so I’m off work. Where would you want to go?

e. Yeah, I should be off work that Monday. Where would you want to go?

f. Yeah, that Monday is Labor Day, so I should be off work. Where would you want to go?

g. Yeah, I might be off work that Monday. Where would you want to go?

h. Yeah, that Monday is Labor Day, so I might be off work. Where would you want to go?
i. Yeah, I’m probably off work that Monday. Where would you want to go?

j. Yeah, that Monday is Labor Day, so I’m probably off work. Where would you want to go?

Again, the ‘must’-claim in (14-a) is marked without further explanation, whereas all the other variants—‘must’ with an argument, as well as non-modal, ‘should’, ‘might’, and ‘probably’—are perfectly acceptable as is.

Two further examples illustrate an important point regarding Support: the argument in question can be shared even if it is not spelled out explicitly, but instead is quietly accommodated. (15) is taken from a radio show:

(15) Mozart wrote the Stadler quintet for his friend Anton Stadler, who must have been a marvelous clarinetist.

The announcer does not give explicit reasons in support of the claim that Anton Stadler was a good clarinetist, but they are easy to recover from the context (the difficulty of the piece, the fact that Mozart wrote it for him).

The second illustration: my phone rings; I can say:

(16) This must be my brother; let me take this.

It is, again, easy for you to recover my reasons for saying this: that I am expecting a call from my brother. (Some subtlety is required here. In general, when someone asserts “Must p”, you can always thereby gather from that some reason to believe p, namely, that the speaker has proposed to update with p. So the fact that Support isn’t always satisfied—as in the infelicitous examples seen so far—suggests that the argument needed to satisfy Support must be more substantive than just that the speaker believes p. I return to this point in the derivation of Support in §3.4.2.)

Support is also evidenced in written forms. Lassiter (2016) discusses a range of examples taken from genealogical discussion boards, noting that ‘Ancestry.com users frequently provide an explicit specification of the evidence used to arrive at a must conclusion.’ Here’s one example Lassiter
discusses, taken from *The Plymouth Colony Archive Project* (slightly altered to improve flow across the variants below):

(17) Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he is not listed among those who were part of the cattle division of 1627—the year we are interested in here; he must have died before then.

A variant which omits the first half of the second sentence, as in (18-a), is felt to be omitting something, if nothing more is said on the subject. By contrast, a non-modal variant with or without that material sounds fine, as do variants with 'might' and 'likely':

(18) a. ??Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he must have died before 1627, the year we are interested in here.

b. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he died before 1627, the year we are interested in here.

c. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he is not listed among those who were part of the cattle division of 1627—the year we are interested in here; he died before then.

d. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he may have died before 1627, the year we are interested in here.

e. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he is not listed among those who were part of the cattle division of 1627—the year we are interested in here; he may have died before then.

f. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he likely died before 1627, the year we are interested in here.

g. Goodman... is listed as one of those who received land in 1623 (PCR 12: 4). However, he is not listed among those who were part of the cattle division of 1627—the year we are interested in here; he likely died before then.

Comparing 'must' with other words that might at first glance seem to work in a similar way, like
'apparently', can help bring out the plausibility of Support. Consider (19), adapted from a television spy drama:

(19)

a. The suspect is fleeing south. We’ve sent agents ahead to Mattapan.

b. Why Mattapan?
   (i) ??The Russians must have a safe-house there.
   (ii) Apparently the Russians have a safe-house there.
   (iii) The Russians have a safe house there.
   (iv) The Russians might have a safe-house there.
   (v) The Russians probably have a safe-house there.

If the conversation ends here, then (19-b-i) is peculiar in a way that the 'apparently', non-modal, 'might', and 'probably' variants are not. 'Apparently' is of particular interest here. 'Apparently', like 'must', is constrained by a form of Indirectness; but 'apparently', unlike 'must', is acceptable here without an argument. Support predicts precisely these contrasts.

Finally, I note that informal polling suggests that the contrasts reported here are robust across strong epistemic necessity modals in English (see §3.4.1 for further discussion of 'can’t) and other languages.

3.2.4 Experimental data

Nearly all the speakers I have polled informally report contrasts for the examples given in the last section exactly where Support predicts them. In this section, I present the results of an experiment which provide independent support for these judgments.

301 Amazon Turk subjects were recruited for the experiment. The stimuli in the experiment comprised the scenario in (14) along with the eight possible responses there, which vary as to

---

11I assume that the indirectness constraint for 'apparently' is somehow lexically encoded.
12Informants report the predicted contrast in Bengali, French, German, Hindi, Japanese, Russian, Spanish, Swiss German, and Turkish.
13I am grateful to Josh Knobe and Jonathan Phillips for providing crucial support in experimental design and analysis for this experiment presented, and to Justin Khoo and the MIT Department of Linguistics and Philosophy for funding for that experiment.
whether there was an argument or not (the argument condition) and as to whether a modal auxiliary was used, and if so, which of three options—‘must’, ‘might’, or ‘should’—was used (the modal auxiliary condition). The scenario and possible responses are repeated here:

Two friends, Scott and Mark, are discussing summer plans. Scott asks Mark: ‘Do you think you’d be free to go fishing in a few months, say in the first week of September?’ Mark responds:

(20) [-arg, +must] Yeah, I must be off work that Monday. Where would you want to go?

(21) [+arg, +must] Yeah, that Monday is Labor Day, so I must be off work. Where would you want to go?

(22) [-arg, -modal] Yeah, I’m off work that Monday. Where would you want to go?

(23) [+arg, -modal] Yeah, that Monday is Labor Day, so I’m off work. Where would you want to go?

(24) [-arg, +should] Yeah, I should be off work that Monday. Where would you want to go?

(25) [+arg, +should] Yeah, that Monday is Labor Day, so I should be off work. Where would you want to go?

(26) [-arg, +might] Yeah, I might be off work that Monday. Where would you want to go?

(27) [+arg, +might] Yeah, that Monday is Labor Day, so I might be off work. Where would you want to go?

Each subject received the scenario, plus one of the eight responses, and then was asked:

(28) Please tell us how natural you think Mark’s response was on the scale below.

The subject responded to this question by dragging a marker on a continuous scale from 0 to 100, with ‘0’ labelled ‘completely unnatural’ and ‘100’ labelled ‘completely natural’.

Each subject saw four scenarios total: one from each modal auxiliary condition, chosen at random within that condition as to whether or not it included an argument. The order of the scenarios
was randomized (thus, for instance, a single subject might have seen [+arg, +should], then [-arg, +must], then [-arg, -modal], then [+arg, +might]). The scenario was designed so that all four modal auxiliary conditions are reasonable to use as a response to Scott’s question, and it was designed so that, in the [-arg] condition, Mark’s follow-up (‘Where would you want to go?’) makes clear that he is changing the subject, and therefore that further information regarding his work schedule will not be forthcoming.

If Support is correct, then subjects should find the [-arg, +must] substantially degraded as compared with all seven other conditions ([+arg, +must], as well as all other modal auxiliary conditions with or without an argument). The responses, which are summarized in Figure 3-1, show precisely this effect: subjects found the [-arg, +must] condition substantially degraded (around a mid-point on the scale of naturalness judgments), as compared with all the conditions.

![Figure 3-1: Mean ratings by condition. Error bars show standard error of the mean.](image-url)
More precisely, analysis of participants’ naturalness ratings revealed a significant interaction between the modal auxiliary condition (‘must’ vs. ‘should’ vs. ‘might’ vs. non-modal) and the argument condition ($F(3, 809) = 11.2, p < 0.001$). To investigate the cause of this interaction effect, I separately analyzed participants’ ratings within each modal auxiliary condition for the [+/-arg] conditions. Within the [-arg] condition, as predicted, participants found Mark’s response with ‘must’ to be substantially degraded—around a mid-point on the scale of naturalness (Mean = 57.20, $SD = 28.86$)—by contrast to all the other modal auxiliary conditions, which they judged around the three-quarters mark on the scale of naturalness ($t’s > -6.08, p’s < .001, d’s > 0.74$).

Things were entirely otherwise in the [+arg] condition, however, where participants found Mark’s response to be quite natural (around the three-quarters mark) in all modal auxiliary conditions, including the ‘must’ condition. Specifically, participants found ‘must’ to be significantly more natural when an argument was included (Mean = 74.55, $SD = 23.51$) as compared to ‘must’ without an argument ($t(257.9) = -5.43, p < .001, d = 0.66$). Participants found the other modal auxiliary conditions with an argument to be roughly equally good, again around the three-quarters mark; participants did not significantly distinguish ‘must’ with an argument from ‘might’ with an argument ($t(267) = -0.51, p = .611, d = 0.06$), though there was some preference for the ‘should’ and non-modal variants over the ‘must’ and ‘might’ variants ($t’s > -3.97, p’s < .001, d’s > 0.48$).

In no condition other than ‘must’ was there a contrast between the [-arg] and [+arg] conditions that was similar to the observed contrast in the ‘must’ condition; in the other modal auxiliary conditions there was some variation between the [+arg] and [-arg] condition, but this variation was small compared with the corresponding variation in judgments in the ‘must’ condition (a +16.99 change in mean naturalness ratings for [+arg, +must] as compared with [-arg, +must], versus a -2.54 change for ‘might’, a +6.25 change for ‘should’, and a +8.75 change for non-modal), and it was not in a uniform direction.

These experimental results confirm the prediction of Support. They show that—as the intuitions elicited throughout this section suggest—‘must’ is felt to be significantly degraded without an argument. By contrast, non-modal claims, ‘should’ claims, and ‘might’ claims are all felt to be relatively acceptable without an argument—around the three-quarters point on the scale of nat-
uralness, as opposed to a mid-point judgment for 'must'. And, as expected, the 'must' variant significantly improved when an argument was added, to around the three-quarters point on the scale of naturalness—roughly the same judgment as for non-modal, ‘might’, and ‘should’ claims, with or without an argument. While there is no doubt much further work to be done to parse the subtleties of judgment here, these results confirm the prediction of Support.

It is worth emphasizing that the results here are graded. Speakers do not reject 'must' statements out of hand when there is no argument, even though they find them to be substantially less good than 'must'-claims with arguments, or other kinds of claims with or without arguments. This is in line with our formulation of Support, and it is an important fact to keep in mind in proceeding. I will ultimately suggest an account of Support as a manner implicature, which I believe is in line with this pattern of judgments. Having said that, it is probably wise not to read too much into this fact, for the following reason. In responding to the stimuli, subjects may be assuming that Scott is able to accommodate an argument for the relevant claim: as we have seen, an argument can become salient to a speaker’s audience through implicit clues, something which it is difficult to control for in the experimental set-up, and which may inflate naturalness judgments in subject’s responses.

3.3 Explaining Indirectness via Support

I thus conclude that Indirectness and Support are both necessary to characterize the difference in felicity conditions between a 'must'-claim and its bare prejacent.\textsuperscript{14} We now turn to the question of how to explain these data. There are three strategies to consider:

1. Account for Indirectness and Support separately.

2. Account for Support in terms of Indirectness, and give an independent account of Indirectness.

3. Account for Indirectness in terms of Support, and give an independent account of Support.

In this section I will pursue the third strategy, after briefly arguing against the first two strategies.

\textsuperscript{14}I will not settle the further question of whether they are jointly sufficient, which is interesting but does not matter for present purposes.
3.3.1 Against strategies 1 and 2

Considerations of theoretical parsimony tell against pursuing the first strategy—providing separate explanations for *Indirectness* and *Support*—unless these separate explanations bottom out in a unified theory. I do not see an attractive way to pursue this strategy, and I therefore set it aside here.

What about the second strategy? This strategy is *prima facie* attractive, since there are a number of extant attempts to give an independent account of *Indirectness*; it is natural to try to recruit them to explain *Support*.

But there are two significant problems with this approach. First, there does not seem to be any way to reduce *Support* to *Indirectness*. A natural first thought is that we can explain *Support* in terms of *Indirectness* by way of a general pragmatic constraint that requires a speaker to share her evidence for a claim if that evidence is indirect. But there is no such pragmatic constraint, as we saw in cases above where non-modal claims were felicitous without shared evidence. A closely related thought is that there is a general pragmatic constraint which requires a speaker to share her evidence if she explicitly indicates the source of her evidence. But, again, as example (19) shows with ‘apparently’—and as cross-linguistic work on evidentials suggests (see Murray 2014)—there is no such constraint: one can use ‘apparently’ or indirect evidential marking without sharing what your evidence is.

A natural second thought is that *Support* reduces to a requirement to assure your interlocutors that *Indirectness* is satisfied. But this approach is not plausible, for a few reasons. First, in most of the cases given above that are felt to be infelicitous without an argument—Patch in her box, Sarah in her windowless office building, the Russian safe-house—there is simply no reason to worry that the speaker’s evidence might not be indirect. Second, it is not generally true that whenever a formulation is constrained by a form of *Indirectness*, the speaker must habitually share her evidence in order to reassure her interlocutors that it satisfies the constraint in question: again, we saw this in (19) with ‘apparently’, which is governed by an *Indirectness* constraint, but which doesn’t require a shared argument. Finally, from a more theoretical standpoint, it is hard to see why an *Indirectness*...
constraint would ever directly yield an obligation to share one’s evidence: we are fairly charitable in assuming that speakers are complying with felicity conditions. For instance, if Indirectness were encoded as a presupposition—on which more in a moment—then, on a standard approach to presuppositions, it will be required that it be common ground that the speaker’s evidence for the prejacent is indirect. But in general interlocutors are perfectly happy to accommodate presuppositions. To derive Support from Indirectness in this way, we would need an explanation of why interlocutors are not in general willing to accommodate an Indirectness presupposition, and instead require that the speaker’s evidence be spelled out so that they can verify that it’s indirect. Perhaps something along these lines can be spelled out, but I do not see how to do it.

This does not close the door to a derivation of Support from Indirectness; but at present I do not see a promising way for this to go.

The second problem with the second strategy is that explaining Indirectness itself has turned out to be quite tricky: both major extant explanations have drawbacks which I will briefly survey here. So it is not clear that, even if we could reduce Support to Indirectness, we would yet have a satisfactory solution to Karttunen’s Problem.

There are two major extant approaches to explaining Indirectness. The first approach is pragmatic. On this approach, an assertion of ‘Must p’ is pragmatically weaker than an assertion of p, in a sense I will make precise in a moment. By choosing the weaker ‘Must p’, rather than the stronger p, the speaker signals that her evidence for p is too weak to merit an assertion of p. Thus, in turn, we can conclude that the speaker’s evidence for p is indirect, since direct evidence would have merited an assertion of p.

This approach seems obviously correct for capturing the correlate to Indirectness for modals like ‘might’, ‘probably’, and epistemic ‘should’ and ‘ought’. Assertions involving one of these (unembedded) modals are uncontroversially weaker than the corresponding non-modal assertions, and thus this account provides a natural explanation of the corresponding indirectness inference.

Nor, in those cases, does a requirement for an argument arise, as we have seen—so there is no

---

17See e.g. Lewis 1979b, Stalnaker 2002 and many others.
18This approach is given in Veltman 1985, Kratzer 1991.
question about how to extend this account of Indirectness to an account of such a requirement. But it is not clear that this approach extends equally well to the Indirectness of ‘must’: the assumption that ‘Must p’ is pragmatically weak is at odds with many intuitions about the relative strength of ‘must’-claims, intuitions which seem better captured instead by Pragmatic Strength:

**Pragmatic Strength**: An assertion of ‘Must p’ makes a proposal which is just as strong as an assertion of p, in the sense that once the common ground is updated with [Must p], it is updated with p.\(^{19}\)

Given the standard assumption that the common ground is meant to track collective behavioral disposition in some sense, Pragmatic Strength essentially says that conversants do not typically leave open the possibility that p is false after accepting ‘Must p’. To see its plausibility, note that (29-b) and (29-c) are decidedly weird responses to (29-a):

(29)  
a. The gardener must be the murderer.

b. ??I concur. Moreover, the gardener is the murderer.

c. ??I concur. Let’s bring him and the butler in to see if we can pin down which of them is the murderer.

Pragmatic Strength directly explains the infelicity of (29-b) as a response to (29-a): if Pragmatic Strength is true, then p adds nothing over and above ‘Must p’. And, likewise, Pragmatic Strength explains the infelicity of (29-c): once a group of interlocutors accepts ‘Must p’, then, if Pragmatic Strength is right, they should no longer leave open the possibility that p is false. By contrast, the infelicity of these sequences is quite difficult to explain if Pragmatic Strength is false. (In §3.5 I argue that there is also a sense in which ‘must’ is weak—a sense which involves less speaker commitment to p when it appears embedded under ‘must’ than when unembedded; my proposal there captures this intuition in a way that is compatible with Pragmatic Strength.)

I will not argue extensively for Pragmatic Strength here; I believe that data like (29) speak strongly in its favor, and, though it has been assumed to be false in pragmatic derivations of Ind-

\(^{19}\) [Must p] is the proposition expressed by ‘Must p’ at context c.
directness, it has not as far as I know been directly challenged in the literature. von Fintel and Gillies (2010) give a battery of arguments in support of a stronger and much more controversial thesis, namely that "Must $p^\neg$ entails that the contextually relevant agents know $p$. This stronger hypothesis entails, but is not entailed by, Pragmatic Strength; Pragmatic Strength is consistent with a weaker semantics on which "Must $p^\neg$ entails $p$ but not knowledge of $p$; or with an account on which "Must $p^\neg$ dynamically entails $p$, in the sense that accepting "Must $p^\neg$ commits you to accepting $p$, but "Must $p^\neg$ does not semantically entail $p$. Lassiter (2016) and Giannakidou and Mari (2016) argue that "Must $p^\neg$ does not entail knowledge of $p$; these data, again, challenge a related, strictly stronger thesis, but do not challenge Pragmatic Strength. But Lassiter presents experimental data which show that subjects agree with "Must $p^\neg$ and $p$ at about equal rates—a prediction in line with Pragmatic Strength, and at odds with its denial; and indeed, though he does not address it explicitly, Lassiter appears to accept Pragmatic Strength, insofar as he adopts a lexical rather than pragmatic derivation of Indirectness (Lassiter remains explicitly agnostic about the stronger hypothesis that "Must $p^\neg$ entails $p$).

If Pragmatic Strength is correct, however, then the present pragmatic derivation of Indirectness does not work.

The main alternative explanation of Indirectness, due to von Fintel and Gillies (2010), posits that epistemic modals lexically encode that the speaker's evidence is indirect. This approach has three relatively theoretical drawbacks which are relatively minor, but are worth noting here. First, as von Fintel and Gillies themselves note, it does not look particularly explanatory, given the robustness of Indirectness for 'must'-claims, both within and across languages. Second, it turns out to be quite difficult to pin down the type of evidence 'must' selects for within standard taxonomies of evidentials, suggesting that categorizing 'must' as a lexical evidential may not be quite right. Third, this approach forces us to take a disunified approach to modality, since other modals (deontic, circumstantial, etc.) do not obviously carry evidential marking.

---

20 And since endorsed and elaborated in Kratzer 2012b, Matthewson 2015, Lassiter 2016.
22 See Sherman 2016 again on this last point, as well as for an alternative approach to Indirectness which aims to avoid these problems. I will not try to evaluate Sherman's approach here, but I note that it insofar as it claims that intuitions
A more serious worry about this approach is that it does not make the right predictions about embedded epistemic modals. On von Fintel and Gillies (2010)'s approach, *Indirectness* is encoded as a lexical presupposition of 'must' that the relevant direct evidence (what they call the kernel) does not *directly settle* the prejacent or its negation: i.e., no single piece of the agent's direct evidence entails the prejacent or its negation.²³ von Fintel and Gillies (2010) observe that negated epistemic possibility modals like 'cannot' carry *Indirectness* to the same extent as 'must', pointing out that, if Billy can see the sunshine, then she can say (30-a), but not (30-b); whereas if Billy has only indirect evidence about the rain, then she can say either (30-a) or (30-b).

(30)  
   a. It's not raining.  
   b. It can't be raining.

To capture this fact within their system, von Fintel and Gillies (2010) propose that 'might' and 'can' have the same lexical presupposition as 'must' (this follows immediately under the assumption that 'might' and 'can' are the duals of 'must').

But this leads to trouble. As Ippolito (2017) points out, 'might'/’can’ do not seem to have an evidential presupposition along these lines. This is clearest when we consider 'might'/’can’ under attitude predicates. On standard theories of presupposition projection, if p presupposes q, then ‘S believes p’ presupposes ‘S believes q’.²⁴ If ‘might’, ‘can’, and ‘must’ carry an indirectness presupposition, then it follows that ‘S believes [must/might/can] p’ will presuppose that S believes that the kernel for the 'must/might/can' does not directly settle p. What is the kernel for an epistemic modal under ‘S believes’? The natural assumption is that it is just S’s direct evidence. This assumption, however, leads to trouble. For instance, since presuppositions project out of questions, (31) is predicted by this approach to presuppose that the police believe their evidence regarding the hair color of the murderer is indirect:

---
²²See Karttunen 1974, Heim 1992 and following. This prediction is fairly uncontroversial, though some have argued that ‘S believes p’ also presupposes q when p does (see Karttunen 1973; Heim 1992; Sudo 2014; Mandelkern 2016). See Geurts 1998 for an argument that ‘S believes p’ only presupposes q, and not ‘S believes q’; we could construct a similar objection if we adopted that theory.
(31) Do the police believe that the murderer might be blond? [modified from (Ippolito, 2017)]

But (31) has no such presupposition. First, note that a response like (32) is perfectly felicitous here:

(32) No, they have ruled it out, based on direct evidence: they’ve caught the murderer and seen that he has black hair.

And, as Ippolito points out, standard responses that target presuppositions, like ‘Hey wait a minute!’, are felt to be out of place here:

(33) #Hey, wait a minute! The police don’t believe that their evidence regarding the murderer’s hair color is indirect.

This response feels like a complete non sequitur.

A possible reply is to treat this as a case of local accommodation (Heim, 1983) where the presupposition is, one way or another, suspended—on a par with cases like (34):

(34) a. Is Susie’s brother coming to the party?
    b. No; she doesn’t have a brother!

But this approach does not seem viable in the present case. First, local accommodation is typically used to explain away counterexamples to otherwise robust patterns. This kind of response will thus be theoretically satisfying only if we find a robust pattern which is occasionally violated (usually with distinctive intonation, timing, and contextual clues). But there does not seem to me to be any inclination, even prima facie, to judge that a indirectness presupposition projects out of ordinary presupposition holes like (31) (more examples to come in a moment).

Second, even when local accommodation is an option, ‘Hey wait a minute’ is generally perfectly felicitous as a response:

(35) a. Is Susie’s brother coming to the party?
    b. Hey wait a minute! Susie doesn’t have a brother!

Ippolito’s example has gender in place of hair color; the switch to hair color makes the subsequent responses more straightforward (since one can obviously directly see hair color, but this is not so obvious with regards to gender).
By contrast, as we have seen, such a response is not felicitous here.

Finally—while a local accommodation story (modulo these drawbacks) could in theory account for the felicity of a negative response like (32)—as Ippolito points out, it would be unable to explain a positive response like any of those in (36):

(36)  
   a. Yes, they believe he might be blond; in fact they know he’s blond, because they have identified the murderer and thus seen his hair color.  
   b. Yes, they believe he might be blond; in fact they believe he is blond, because they have identified the murderer and thus seen his hair color.  
   c. Yes—they’ve seen that he is blond!

Whether or not the presupposition of (31) were locally accommodated, an affirmative reply to (31) would endorse that presupposition, and so any of the responses in (36) should entail that the police believe their evidence about the murderer’s hair color is indirect. These responses thus should be felt to be contradictory—if the police know that the murderer is blond based on having identified him and seen his hair color, then their evidence that he is blond is direct, in any relevant sense, and so these responses entail both that their evidence is direct (as part of their asserted content) and that it is indirect (thanks to the presupposition they affirm). But the responses in (36) are not felt to be contradictory.

The patterns observed here regarding projection out of questions are robust across presupposition “holes”, like the scope of negation, modals, and the antecedents of conditionals:

(37)  
   John doesn’t believe that the keys might be in the car [. . . because he sees them on the table].  
   [modified from (Ippolito, 2017)]

(38)  
   John may think the keys might be in the car. [. . . But he also might see that they’re sitting on the table.]

(39)  
   If John thinks the keys might be in the car, then he’ll look there. [. . . But he might already see that they’re sitting on the table.]
All the points regarding (31) apply here: on a presuppositional approach to Indirectness, these variations should all presuppose that John thinks his evidence with regard to the location of the keys is indirect. But this does not seem to be an inference licensed by these assertions. This is suggested by the same kinds of evidence we saw above: first, the felicity of the bracketed possible continuations (which conflict with this inference), and second, the infelicity of ‘Hey wait a minute’ responses. And, as above, appealing to a local accommodation story does not look appealing, since ‘Hey wait a minute’ is not an appropriate response here.

There is room for maneuver here by relaxing the assumption that the kernel for ‘must’ under ‘S believes’ is supplied by S’s direct evidence.26 But whose direct evidence should it be instead? It is not clear there are any good options. Suppose that the relevant direct evidence is the evidence of some individual or group $x$. Then a sentence like (37) should presuppose that John thinks that $x$’s direct evidence is indirect. Now suppose John has the Truman Show delusion: he thinks that everyone on earth is watching his every action, and can see everything she sees. Moreover, let’s say that he thinks this is common knowledge among everyone on earth. Then the bracketed continuation should be infelicitous, since, if John saw the keys on the table (and is aware that he did), then he will believe that, for any group or individual $x$, $x$ also saw them on the table, and so he will not believe that anyone’s evidence regarding the location of the keys is indirect. Thus it is not clear that going down this path helps: no matter how we interpret modals under attitudes, it looks like a presuppositional approach will predict projection patterns which are not in fact observed. It thus does not seem promising to me to treat Indirectness as a presupposition.

### 3.3.2 From Support to Indirectness

Both parts of the second strategy—reducing Support to Indirectness and then giving an independent explanation of Indirectness—thus face substantial open challenges. By contrast, I will argue now that the third strategy—deriving Indirectness from Support, and then giving an independent explanation of the latter—provides an empirically and explanatorily satisfying approach.

In brief, I will argue that we can derive Indirectness from Support as follows. First, Pragmatic

---

26Following broadly the strategy in von Fintel and Gillies 2011. Thanks to Kai von Fintel for discussion.
Strength says that an assertion of ‘Must p’ is a bid to update the common ground with p. Support says that it is a proposal to do so on the basis of an argument Γ. General principles forbidding redundant assertions entail that p should not follow from Γ in a way that is mutually recognized to be obvious. Finally, speakers are generally obligated to give their best argument for p if they’re giving an argument for p at all. It follows that, in order for an assertion of ‘Must p’ to be felicitous, p should not follow in a mutually obvious way from the best argument a speaker of ‘Must p’ can have for p. In short, an assertion of ‘Must p’ is a proposal to accept p on the basis of a shared argument; this is the kind of proposal one should make only when that argument leaves some epistemic space between its premises and conclusion.

The first step in our derivation is the assumption, briefly defended in §3.3.1 above, that an assertion of ‘Must p’ is pragmatically strong in the sense that it is just as strong as an assertion of p: i.e. that an assertion of ‘Must p’ is, inter alia, a proposal to update the common ground with p.

The second step is to note that in general, when a speaker gives an argument in support of p with the intention of getting her interlocutors to accept p on the basis of that argument, the argument in favor of p must be non-redundant in some sense. Compare the two variants in each of (40) and (41):

(40)  a. I put Patch in her box this morning, and no one has let her out. So she’s in her box.
     b. ??I see Patch in her box. So she’s in her box.

(41)  a. Clinton has amassed a majority of pledged delegates and superdelegates. So a woman will clinch the Democratic nomination!
     b. ??Clinton will clinch the Democratic nomination. So a woman will clinch the Democratic nomination!

(40-b) strikes me as objectionable; there is something pedantic or redundant about it. Likewise for (41-b). By contrast, (40-a) and (41-a) are fine. The difference seems to be that in (40-a) and (41-a), there is enough epistemic space left between the argument in the first sentence and its conclusion in the second that its conclusion is not felt to be redundant. This intuition can be regimented as a norm against redundant assertions, along the following lines:
Non-Redundancy: A proposal to update the common ground with \( p \) on the basis of an argument \( \Gamma \) is infelicitous if \( p \) follows from \( \Gamma \) in a way that is mutually recognized to be obvious.

In the next section I will further explore the justification for Non-Redundancy, show that it follows from close variants on widely accepted norms, and say more about what it amounts to for \( p \) to follow from \( \Gamma \) in a way mutually recognized to be obvious. Note for the present, however, that Non-Redundancy nicely captures the contrast between (40-a) and (40-b). The first is acceptable, since there is some epistemic space between the premises—having put Patch in her box in the morning, together with no one else having let her out—and the conclusion—that Patch is in the box (epistemic space filled by background assumptions about Patch’s ability to get out on her own, etc.). The second is not, since it does follow in a mutually obvious way from seeing Patch in her box that she is in her box. Similar considerations apply to the contrast between (41-a) and (41-b).

The last step in our derivation says that a speaker must give the best argument for \( p \) that she has, if she’s giving an argument for \( p \) at all. To see the plausibility of this constraint, consider (42):

(42)  [John was at the Red Sox game and knows on this basis who won. He also read about the game in the Boston Globe.]

a.  [Max:] Who won the game?

b.  [John:] ?? The Red Sox, according to the Globe.

If John intends (42-b) to answer Max’s question, then there is something strange about it; we expect John to give his strongest evidence for the claim that the Red Sox won. In general, speakers are required to share the best piece of evidence they have for a claim, if they are sharing evidence at all. This follows naturally from a broadly Gricean vantage point on conversational dynamics: in (42-b), John is violating Grice’s Maxim of Quantity by failing to ‘make his contribution as informative as is required (for the current purposes of the exchange)’ (Grice, 1989). More precisely, the lesson of

\(^{27}\)Note that Non-Redundancy does not forbid post hoc support for an assertion with a redundant argument; it is perfectly fine to justify oneself, if challenged, with ‘Because I saw it’. What Non-Redundancy forbids is making an initial bid to update the common ground with something on the basis of an argument from which it follows in a mutually obvious way.
cases like this is a corollary of the Maxim of Quantity which I call Strongest Evidence:28

**Strongest Evidence**: When a speaker aims to update the common ground with p on the basis of an argument Γ, she is obligated to do so by providing the strongest argument—the best piece of evidence—which she has for that claim.

We can now put these pieces together to derive Indirectness from Support. Support says that an assertion of "Must p" is felicitous only if there is a shared argument for p. Pragmatic Strength says that an assertion of "Must p" is a proposal to update the common ground with p. I will make the plausible further assumption that an assertion of "Must p" is thus a proposal to update the common ground with p on the basis of a shared argument for p (this is an assumption that will fall out of the derivation of Support below). According to Non-Redundancy, p must not follow from that argument in a mutually obvious way. According to Strongest Evidence, that argument must constitute the best evidence the speaker has for p. It follows that in order for a speaker to be able to felicitously assert "Must p", p cannot follow in a mutually obvious way from the speaker's best piece of evidence for p. In other words, the speaker's best evidence for p must be indirect, in whatever sense of indirectness is relevant to evaluating whether an argument is felt to be redundant.

Put differently: if a speaker has direct evidence (in the sense relevant to judgments about redundancy) for p, then, if she were to assert "Must p", then due to Strongest Evidence and Support, she would have to give that evidence as an argument on the basis of which she is proposing her interlocutors accept p; but then she would be bound to violate Non-Redundancy. So if she has direct evidence for p, she cannot assert "Must p".

In sum: in asserting "Must p", the speaker has to ensure there is a shared argument which represents her best evidence for p, and yet is not so strong that it makes the 'must'-claim sound redundant. Thus p can't follow in a mutually obvious way from her best evidence for p. No parallel constraint follows for non-modal claims—since Support requires only that 'must'-claims be supported by an argument—and thus Support, plus Pragmatic Strength, Non-Redundancy, and Strongest Evidence,

28See Faller 2012 for more careful discussion of how this kind of reasoning would go. To spell out Strongest Evidence in more detail, we need to be able to access a scale of evidential strength, according to which, say, direct perceptual evidence counts as stronger than any kind of testimonial evidence—on this point see also Faller 2001. A norm like Strongest Evidence is related to but orthogonal to knowledge norms on assertion of the kind discussed e.g. in Williamson 2000.
entail a form of *Indirectness*.

### 3.3.3 Non-Redundancy

This completes my derivation of *Indirectness* from *Support*. In the next section I will turn to the question of how to predict *Support*. Before doing so, however, I will say more about the conversational architecture that underlies *Non-Redundancy* (in this subsection), as well as the predictions made by the present derivation of *Indirectness* (in the next subsection).

Much more could be said about each of the assumptions I made in the last section, but *Non-Redundancy* is the one most in need of further explanation here, since it is a central upshot of my account that the signal of indirectness associated with 'must' reduces to judgments about redundancy. Saying more about redundancy will help us evaluate this claim.

*Non-Redundancy*, again, runs as follows:

*Non-Redundancy*: A proposal to update the common ground with \( p \) on the basis of an argument \( \Gamma \) is infelicitous if \( p \) follows from \( \Gamma \) in a way mutually recognized to be obvious.

*Non-Redundancy* follows from two premises. First, proposing to update the common ground with \( p \) on the basis of an argument \( \Gamma \) just is a proposal to update the common ground with \( \Gamma \) and then with \( p \). Second, one should not propose a series of updates if, should they all be accepted, the final update will be judged to be redundant.

To spell out this simple idea, we need to investigate more general principles forbidding redundant assertions. The commonly accepted principle forbidding redundant assertions is typically formulated along the following lines:

*Common Ground Entailment*: Don’t assert \( p \) if \( p \) is entailed by the common ground.\(^{29}\)

This norm is very natural if we think of conversations as cooperative enterprises whose goal is information transfer. Given our limited cognitive and temporal resources, we should not say what is

---

\(^{29}\)See Stalnaker 1974, 1978, 1973. Some version of this norm applies at the subsentential level as well; see e.g. Schlenker 2008a, 2009, Mayr and Romoli 2016.
already common ground. *Common Ground Entailment*, however, is not quite right. It overgenerates, for instance, in cases like (43):

(43)  
   a. If it’s raining out, then Bob brought his umbrella. If Bob brought his umbrella, then he won’t have noticed that we had the roof redone. And, it was raining out. 
   b. Ok. 
   c. So Bob won’t have noticed that we had the roof redone. 

*Common Ground Entailment* wrongly predicts that (43-c) will be infelicitous. (43-c) is acceptable, however: it simply makes explicit the conclusion of a somewhat complex chain of reasoning, one which the speaker would not necessarily have expected her interlocutor to draw on her own. The problem, intuitively, is that *entailment* is too strong of a notion to play the role it does in *Common Ground Entailment*. Instead, I propose to treat an assertion as redundant if it is entailed by the common ground in a way mutually recognized to be obvious, in a derogatory sense: a sense which makes an assertion superfluous. In the present case, we say that (43-c) is not a mutually obvious entailment of (43-a), and thus we rightly predict its felicity. What counts as a mutually obvious entailment will depend on the context, and will be a matter of ongoing negotiation; I discuss this point further below.  

Merely changing ‘entailed’ to ‘entailed in a mutually obvious way’ in *Common Ground Entailment* does not yet, however, yield a satisfying principle. The modified principle will undergenerate in important cases. Consider:

(44)  
   a. What time is the movie? 
   b. The cinema website says that it’s at 7:30. 
   c. Ok. 
   d. ??So the movie’s at 7:30. 

(44-d) is, in most contexts, unacceptable—intuitively, because it is felt to be redundant. But it is not

---

30See Barker 2009 and Kripke 2009 for related discussion, as well as literature on the normativity of logic (e.g. MacFarlane 2004). As Justin Khoo has pointed out to me, this will help explain why it can be rude to spell things out in detail: this shows that you are not treating something as mutually obvious which your interlocutor assumed was.
as though (44-b) *entails* (44-d). To capture data like this, we need to use a more permissive relation than entailment: a relation which can capture the sense in which ‘the movie is at 7:30’ follows from ‘the cinema website says that the movie is at 7:30’.

Here I will simply use ‘follows from’ to capture this relation, with the understanding that we will get a feel for the notion from examples like (44). If \( \Gamma \) entails \( p \), then \( p \) follows from \( \Gamma \); but \( p \) also might follow from \( \Gamma \) if \( p \) is a reasonable default inference, or is highly probabilified by \( \Gamma \). Much more needs to be said here, but this suffices for our purpose.

The most natural way to capture these two revisions to Common Ground Entailment is to model the common ground as a set of propositions which is closed under mutually obvious inference— if \( p \) is in the common ground and \( q \) follows from \( p \) in a mutually obvious way, then \( q \) is in the common ground—but which is not closed under logical entailment in general. Then we can state our redundancy principle as follows:

**Common Ground Settlement:** Don’t assert \( p \) if \( p \) is in the common ground.

Note that if we model the common ground this way, then—if we follow standard assumptions in treating it as derivative from individual mental states—the present considerations can be extended to argue in favor of likewise modeling individual attitudes as sets of propositions with these closure properties.

There is, again, much more to be said about Common Ground Settlement—about when something follows from a set of premises, when it does so in a way mutually recognized to be obvious, and what the underlying theory of mind should be to account for this. I leave further exploration of these topics for future work. What is important for present purposes is that Non-Redundancy is just a corollary of Common Ground Settlement. If a speaker proposes to update the common ground

---

31 A different response to data like this would be to hold that in cases like this, we accommodate an enthymematic premise, so that this really is an entailment. I do not see how to choose between this response and the present one; either formulation would do for our purposes.

32 See e.g. Stalnaker 2002.

33 The appropriate semantics for attitude reports would thus be a neighborhood semantics, rather than a standard modal semantics (Montague 1970, Scott 1970). See Kratzer 2012c, 19-20 for a general defense of representing attitudes with sets of propositions. An alternate approach would be to model the common ground, and thus mental states in general, as overlaid with a partition (see Roberts 2012; Yalcin 2011, 2016; Bledin and Rawlins 2016). This approach accomplishes much the same goal as the present one but is more constrained than the present approach.
with \( p \) on the basis of an argument \( \Gamma \), then she intends her interlocutors to first accept the elements of \( \Gamma \) and then—on that basis—accept \( p \). In other words, she intends the elements of \( \Gamma \) to become common ground, and then for \( p \) to become common ground. In order for an action of this kind to be acceptable, according to *Common Ground Settlement*, it must be the case that \( p \) does not follow from \( \Gamma \) in a way mutually recognized to be obvious.

### 3.3.4 Predictions

Having spelled out *Non-Redundancy* more carefully, we are in a position to flesh out a striking empirical prediction of our account: namely, that S’s evidence \( \Gamma \) for \( p \) counts as indirect in the sense relevant to *Indirectness* just in case an assertion of \( p \) following sequential assertions of the elements of \( \Gamma \) does not strike us as redundant. This provides a principled characterization of the notion of indirectness involved in *Indirectness*, a characterization which I will argue nicely accounts for our judgments about ‘must’-claims.

I will highlight a few points in this respect. First, as noted above, von Fintel and Gillies (2010) observe that ‘must’-claims based on reliable testimony are generally not acceptable:

\[
(45) \quad \text{The website says the movie is at 7:30. So the movie must at 7:30.}
\]

That is, reliable testimony counts as indirect in the sense relevant for assessing ‘must’-claims. This follows immediately in our approach. This is because reliable testimony for \( p \) is typically felt to be a redundant argument for \( p \), as shown by examples like (44) (a non-modal variant of (45)). (*Why* sequences like this are treated as redundant is, of course, an important further question for theories of redundancy to address, but one I will not answer here.)

The second prediction of our approach worth highlighting is that what counts as redundant in a given context—and thus judgments about the felicity of ‘must’—will be context-sensitive, since *Non-Redundancy* is context-sensitive. In particular, what counts as redundant in a context depends on what counts as mutually obvious in that context. Thus, e.g., while (45) is infelicitous out of the blue, it may be felicitous in a context in which the inference from website listings to fact is not generally accepted, as in (46):

(46) ??The website says the movie is at 7:30. So the movie must at 7:30.
Google says that the movie is at 7:30. Websites listing movie times are generally extremely unreliable. Google is extremely reliable, though, so the movie is indeed at 7:30.

If *Indirectness* bottoms out in judgments about redundancy, we thus predict that what counts as indirect will, likewise, be a context-sensitive matter. This prediction, again, is borne out. While (45) was strange out of the blue, it is much improved in (47), which, like (46), calls into question the reliability of websites listing movie times:

(47) Google says that the movie is at 7:30. Websites listing movie times are generally extremely unreliable. Google is extremely reliable, though, so the movie must indeed be at 7:30.

Third, we can explain why ‘must’-claims that conclude a complicated argument are generally acceptable, even if the premises of the argument entail its conclusion. ‘Must’ is often warranted in mathematical or logical contexts, like (48).34

(48) If the set of validities were decidable, then the halting problem would be decidable. The halting problem is not decidable. So the set of validities must be undecidable.

Our approach makes sense of this in a straightforward way, since the conclusion of (48) does not follow in a way that is mutually obvious from the premises (even if it is entailed by them, and indeed by anything).

A final set of predictions of the present account has to do with the structure of my derivation of *Indirectness*. I have proposed that *Indirectness* arises due to conversational norms. It is a hallmark of pragmatic phenomena like this that they can be cancelled, since the underlying conversational norms are generally defeasible. We thus predict that *Indirectness* will be cancelled when one of the underlying norms is not in play. This prediction, again, is borne out. Consider first contexts in which **Strongest Evidence** is not in play because it is overridden by considerations which prevent

---

34 Is the ‘must’ here epistemic? Some have argued that this ‘logical’ ‘must’ is not genuinely epistemic (e.g. Giannakidou and Mari 2016, Goodhue 2016). While I cannot say anything to decisively rule out that possibility, I note two things that militate against this option and in favor of a unified approach: first, it is inelegant to multiply modal flavors further than we need to. Second, even if we say that the logical ‘must’ is not epistemic, we still need a theory of its distribution, since it is not always warranted, even when its complement is a logical consequence of the common ground; simply saying that this ‘must’ is logical, not epistemic, thus does not yet explain its behavior. My account avoids multiplying categories of modality, and explains the distribution of ‘must’ in this kind of environment.
the speaker from sharing her strongest evidence for \( p \). For instance, suppose that Mary is at Tom’s party. She goes out to the street to smoke, where she runs into Ben. She knows Ben wasn’t invited to the party, and doesn’t want him to know that she was invited. Ben can hear music coming from Tom’s place, and asks Mary what’s going on at Tom’s. Mary wants to communicate that he’s having a party, but she doesn’t want to share her strongest evidence for this—and doesn’t seem to be under any obligation to do so, since she is trying not to hurt Ben's feelings. In this context, she can felicitously assert (49):

(49) Given the music, it must be some kind of party.

(49) may be misleading, but it strikes me as perfectly felicitous, despite the fact that Mary has direct evidence that there is a party. This is precisely our prediction, since \( \text{Strongest Evidence} \) seems to be suspended here.

It is somewhat harder to find similar cases in which \( \text{Non-Redundancy} \) is suspended, since \( \text{Non-Redundancy} \) already has an element of context-sensitivity built in (since what counts as 'mutually obvious' in the relevant sense is context-dependent). But if we can find contexts in which it is suspended, then we predict that in those contexts as well, \( \text{Indirectness} \) will be suspended. One kind of context might be when a speaker wishes to signal that her interlocutor is being dense, and thus spells out reasoning even when it is mutually obvious:

(50) a. I can’t believe it. Are you really firing me?
   b. Well, let’s see. I’m your boss. Under your contract, I have complete discretion to terminate your employment at any time. I told you in writing that you’re fired. So, yep, I’m firing you.

In contexts like this, the boss is spelling out reasoning that is mutually obvious; her assertion seems to be felicitous, if obnoxious, and so \( \text{Non-Redundancy} \) seems to be suspended here. We thus predict that \( \text{Indirectness} \) will be suspended as well—that is, that ‘must’ is perfectly felicitous here, even though the boss’s evidence about whether she is firing her employee is perfectly direct. This seems to be borne out: (51) is felicitous (if, again, obnoxious):

(51)
I can’t believe it. Are you really firing me?

Well, let’s see. I’m your boss. Under your contract, I have complete discretion to terminate your employment at any time. I told you in writing that you’re fired. So, yep, I must be firing you.

These points suggest that, as I have argued, Indirectness is indeed a pragmatic inference, which will therefore be suspended when the underlying pragmatic principles are suspended. More generally, the discussion in this section shows that deriving Indirectness in the way I have proposed allows us to make sense of the subtle pattern of judgments about the felicity of ‘must’-claims.

3.4 Support

Support plus independently motivated pragmatic principles thus provide a theoretically and empirically satisfying explanation of Indirectness. I turn now to the question of how to account for Support. I briefly survey and criticize the few extant proposals before sketching my own proposal. I should note here that—just as my account of Indirectness in the last section was separable from my account of the empirical picture—likewise my account of Support in this section is separable from my account of Indirectness, as well as my account of the empirical picture.

3.4.1 Extant proposals

Support says, roughly that ‘must’ requires that an argument for its prejacent be made salient (and that no similar requirement applies to other modal or non-modal claims). A natural first thought about how to account for Support is to treat ‘must’ as containing something like an implicit indexical which refers to an argument: ‘Must p’ means roughly ‘This argument entails p’, where the implicit ‘this’ requires a salient referent. Stone (1994) suggests an account along just these lines: on his approach, ‘must’ has a lexical argument place which needs to be saturated by an argument made salient by context. In other words, ‘must’ is a two place operator, taking as arguments a sentence p and an argument Γ, which says that Γ provides decisive reason to believe p.

A solution along these lines, natural though it is, does not work. The issue is that, as we
saw above, 'might' and 'can', unlike 'must', do not require a salient argument. We saw abundant evidence for this in §3.2; to remind you of the contrast, consider (52):

(52) [Julie's cat has been sneezing a lot. Ben asks her how the cat is doing. Julie says:]

a. Not so great. I need to take him to the vet actually, he might have an upper respiratory infection.

b. Not so great. I need to take him to the vet actually, he has an upper respiratory infection.

c. ??Not so great. I need to take him to the vet actually, he must have an upper respiratory infection.

Suppose the conversation ends here. As Support predicts, the 'must' variant in (52-c) is infelicitous as it stands, without an argument; a response like (52-c) only works if Julie follows it up with something like 'He's been sneezing a lot lately'. By contrast, the 'might' variant in (52-a), like the non-modal variant in (52-b), is acceptable without an argument. In short: 'might', unlike 'must' but like non-modal claims and other kinds of modal claims, does not require an argument for felicity.

If we took Stone's approach, however, then, assuming that 'must' and 'might'/'can' are duals, we would predict that 'might' has an anaphoric requirement for an argument, just as 'must' does: if 'might' means 'not must not', then the argument requirement of 'must' will project through negation, and thus 'might' will require a salient argument, too.

We could avoid this by giving up the assumption that 'must' and 'might'/'can' are duals, and that 'might'/'can' does not have a lexical argument place for an argument. But going this way would lead to a serious new puzzle. Assuming we treat 'cannot' as 'not (can)', then, if we go this way, we will predict that 'cannot', like 'can', does not have an anaphoric requirement for an argument any more than (unembedded) 'can' does. But this is wrong: the same examples we used to motivate Support for 'must' above can be used to motivate it for unembedded 'cannot' (modulo obvious changes). Thus, for instance, consider (53):

(53) Emma notices that her neighbor Phil hasn't taken in his mail in some time, and concludes
that he is out of town. Another neighbor asks if Phil is around. Emma responds:

a. ??No, he can’t be.
b. No, he’s not.
c. No, he can’t be: no one has taken his mail in for a week.
d. No, he’s not: no one has taken his mail in for a week.

The exchange ends here.

Just as for ‘must’, and just as predicted by Support (since ‘can’t’ is a strong epistemic necessity modal), the ‘can’t’ variant without an argument in (53-a) is marked, as compared with ‘can’t’ plus an argument, or non-modal variants with or without an argument. As for ‘must’, then, ‘can’t’ requires that an argument be made salient. If we treat ‘might’ and ‘can’ as lacking a lexical requirement for an argument, then we will also predict that ‘can’t’ lacks a lexical requirement for an argument, and thus we will fail to make sense of the fact that ‘can’t’ requires an argument.

Thus a lexical stipulation of Support along the lines Stone suggests faces a dilemma: either treat ‘must’ and ‘might’/‘can’ as duals, and wrongly predict that the latter require a salient argument; or do not treat them as duals, and wrongly predict that ‘cannot’ does not require a salient argument.35

Neither option is acceptable, and so I conclude that an approach like this does not work.

35One might at this point try to claim that ‘can’ and ‘might’ in fact differ in just the respect that ‘can’ has an argument place for an argument, while ‘might’ does not. This claim is hard to evaluate in English, since ‘can’ is rarely interpreted epistemically when it is not negated. However, it is fairly easy to dismiss this proposal when we look at an epistemic modal like ‘could’ in English, or at an epistemic modal in a language like German, both of which can be embedded under wide-scope negation:

(54) The murderer could be the gardener.
(55) The murderer couldn’t be the gardener.
(56) Der Mörder könnte der Gärtner sein.
The murderer can the gardener be.
‘The murderer might be the gardener’.
(57) Der Mörder könnte nicht der Gärtner sein.
The murderer can not the gardener be.
‘The murderer can’t be the gardener’.

(55) and (57), but not (54) or (56), require a salient argument.
Similar criticisms extend to the account suggested in Swanson 2015, who builds on Kratzer 1981 in adopting a premise semantics for epistemic modals, with the added requirement that those premises be publicly available.\(^{36}\) This approach will again face a dilemma. If we treat ‘might’/‘can’ as the duals of ‘must’, then this explanation will overgenerate: it will wrongly predict that they are likewise subject to a requirement for a salient argument, since they will likewise require a set of premises to be made public. Alternatively, we could abandon duality, but then we cannot explain the requirement of ‘cannot’ to have a salient argument.\(^{37}\)

This kind of objection also tells against a treatment of Support as a presupposition.\(^{38}\) Presuppositions project through negation; thus a presuppositional approach would either treat ‘might’/‘can’ as duals of ‘must’, and thus wrongly predict that ‘might’/‘can’ require a salient argument; or it would abandon duality, and once again fail to predict that ‘cannot’ requires a salient argument.

### 3.4.2 Support as a manner implicature

We can avoid these problems by deriving Support as a manner implicature. The derivation depends on a key assumption about the interpretation of unembedded epistemic modals: namely, that their interpretation makes reference to the common evidence of the group of interlocutors. I should emphasize at the outset that this is not a commitment about the semantic value of epistemic modals (and therefore, importantly, does not bear on the interpretation of embedded modals); and so this assumption is compatible with a variety of semantics for epistemic modals, in particular the standard quantificational semantics of Kratzer 1977, 1981, as well as the semantics spelled out in Chapter 1. On both these approaches, an unembedded assertion of ‘Must p’ is interpreted as meaning, roughly, ‘p is entailed by the contextually relevant information’\(^7\).\(^{39}\) The present claim can then be interpreted as the claim that the contextually relevant information is typically determined by the

---

\(^{36}\)Moss (2015) gives a proposal which is relevantly similar to Swanson’s for our purposes.

\(^{37}\)An alternative, suggested to me by Eric Swanson (p.c.), maintains duality, but treats the salience of the argument in question as part of the asserted content of ‘must’: ‘Must p’ means ‘There is a salient argument Γ which entails p’; and so ‘Might p’ means ‘There is no salient argument Γ which entails not p’. While this does avoid the present objection, it seems to me to make the wrong prediction about the meaning of ‘might’; it predicts that the absence of any salient arguments entails the truth of all ‘might’ claims. Suppose that I know p, but no argument for p is salient. I nonetheless cannot assert ‘Might p’, even though its truth conditions, on this account, would obtain in this scenario.

\(^{38}\)A suggestion due to Eric Swanson (p.c.).

\(^{39}\)Ignoring the role of a normality ordering source in Kratzer’s semantics. I will assume a relational framework here, though in theory, we could couple the present assumption with different modal semantics.
common evidence of the group of interlocutors. This kind of approach is closely related to the approach spelled out in Chapter 2; as we saw there, this approach provides a way to make sense of the basic update effects of modal claims, in particular the way they are used to coordinate on the structural properties of the common ground. Spelled out in an appropriate way, this approach predicts that an assertion of \( \Box \text{Might } p \) is felt to be a proposal to make \( p \) compatible with the common ground; and—crucially for our purposes—it predicts that an assertion of \( \Box \text{Must } p \) is felt to be a proposal to make \( p \) entailed by the common ground, since, if it is common ground that the common evidence entails \( p \), then, given plausible assumptions about the relation between evidence and belief, \( p \) itself will be common ground. This is in line with Pragmatic Strength—an assertion of \( \Box \text{Must } p \), if accepted, has the effect of adding \( p \) to the common ground—without committing to the stronger, and more controversial, hypotheses mentioned above (that \( \Box \text{Must } p \) entails \( p \), or entails knowledge of \( p \)).

On this approach, an assertion of \( \Box \text{Must } p \) thus has the same basic update effect as an assertion of \( p \) alone: if a group accepts either one, then they accept \( p \). Note, moreover, that \( p \) is structurally simpler than \( \Box \text{Must } p \). Since they have the same basic update effect, they will therefore be in competition: if a speaker chooses a more complex way to achieve the same basic effect, this choice will require some explanation. What could explain this choice? Note that in the present framework, \( \Box \text{Must } p \) and \( p \) have different subject matters. \( p \) on its own is simply about whether \( p \) is true. \( \Box \text{Must } p \), by contrast, is about whether \( p \) follows from the interlocutors’ common evidence. An assertion of \( \Box \text{Must } p \) is thus like an assertion of \( p \), in that it is a proposal to accept \( p \) which goes by way of calling

---

40In particular, with the conversation’s common evidence after the assertion in question is made.
41In particular, this will follow if we assume that, if it is common ground that \( p \) is entailed by the common evidence, then \( p \) will be common ground; and if it is common ground that \( p \) is compatible with the common evidence, then \( p \) will be compatible with the common ground. In that case, this kind of view will be essentially equivalent to the view spelled out in Chapter 2.
42Though these would follow if we treat evidence as a factive notion.
43I will not spell out assumptions about how we calculate which alternatives are relevant. It seems fairly plausible that however we do so, \( p \) will count as a relevant alternative to \( \Box \text{Must } p \) (and \( \Box \text{Not } p \) as a relevant alternative to \( \Box \text{Can’t } p \)); this follows e.g. on the account given in Katzir 2007, according to which alternatives are calculated by the deletion or replacement of nodes at LF.
44There are a variety of ways of spelling out this intuition; it doesn’t particularly matter for our purposes how to do so. A simple approach just says that the subject matter of an assertion of \( p \)—the question raised by such an assertion—is simply the question \( \{ p, \neg p \} \). See Lewis 1988, van Kuppevelt 1995, Ginzburg 1995a,b, Roberts 2012, Yablo 2014, Bledin and Rawlins 2016 and citations therein.
attention to the interlocutors' common evidence for p.

The choice between "Must p" and p is thus a choice between proposing an update with p which calls attention to the interlocutors' evidence for p, versus proposing an update with p which does not. Asserting "Must p" is thus a way to propose an update with p which draws attention to the interlocutors' evidence for p. Given that the speaker had a strictly less complex alternative which does not do so, an assertion of "Must p" will thus be felt to be a proposal for her interlocutors to accept p with special attention to their evidence for p.

What kind of evidence would be worth drawing attention to in this way? One possibility is that the evidence is just that the speaker herself endorses p. But there is nothing interesting or noteworthy about this kind of evidence: this kind of endorsement is in play whenever the speaker proposes an update with p—including when she asserts p alone—and thus, if this were the case, the speaker would have no reason not to assert p alone, the structurally simpler alternative. Calling attention to the speakers' evidence for p will thus make sense only if the speaker wants her interlocutors to accept p based on substantive evidence for p—evidence that goes beyond the speaker's endorsement of p. And if the speaker wants her interlocutors to accept p based on evidence like this, then she will, of course, need to ensure that such evidence is available to them—either by providing it, or by ensuring that her interlocutors can recover it from the common ground (possibly by accommodation).

In short: if a speaker chooses to achieve the same basic update effect with a strictly more complex update, interlocutors will ask why she chose the more complex option. If the only relevant difference between the two is that the more complex option calls attention to the interlocutors' common evidence for p, the interlocutors will conclude that the speaker finds this evidence noteworthy: she wants her interlocutors to reason to the conclusion that p is true on the basis of that evidence, rather than on the basis of her epistemic authority. And so, of course, she needs to be sure that there is common evidence for p, explaining why 'must' is felt to be degraded if there is no argument made salient (or recoverable from context).\textsuperscript{45}

\textsuperscript{45}von Fintel and Gillies (2010) cast doubt on a pragmatic derivation of \textit{Indirectness} which bears some similarity to the present approach: it treats "Must p" and p as competitors, with the latter stronger than the former. The problem, of course, as we have seen, is that the latter is not stronger than the former. Like von Fintel and Gillies, I am giving a pragmatic derivation of \textit{Support} which treats "Must p" and p as competitors. Crucially, though, we avoid the analogue to von Fintel and Gillies' objection, because this derivation does not rely on the assumption that one of these is stronger than the other: instead, it goes by way of considerations of complexity and the subject matter of 'must'-claims.
Importantly, this approach to Support avoids the problems sketched for existing views above. In particular, this approach predicts that only strong epistemic necessity modals will require an argument. The reason for this is that a key step in the reasoning above was the assumption that "Must p" and p alone are in competition. That, in turn, follows because of the assumption that "Must p" and p alone have the same basic update effect (namely, adding p to the common ground). This assumption will be merited for all strong epistemic necessity modals, but only for strong epistemic necessity modals: weaker epistemic modal claims like "Might p", "Probably p", "Ought p", and so on uncontroversially have a weaker update effect than p alone: none of these are felt to be proposals to add p to the common ground. And so—unlike existing approaches, which, as we saw above, either wrongly predict that "Might p" requires an argument, or that "Can't p" does not—we rightly predict that the reasoning just sketched will be blocked for these weaker modals, and thus that we will find the characteristic pattern embodied in Support: strong epistemic necessity claims require an argument, while weaker epistemic modal claims, like non-modal claims, do not.\(^{46}\)

Is Support ever cancelled in the way that, as we saw above, Indirectness can be? It is not clear to me that it is. The pragmatic principle underlying the derivation is something like "If p and q are alternatives with the same basic update effect, use the structurally simpler one unless you have a good reason not to". It is not clear that this principle is ever suspended, and thus not clear to me that we predict that Support is ever cancelled. On the other hand, a derivation of Support as a manner implicature predicts that a violation of Support—failing to provide an argument with a ‘must’-claim—will be relatively innocuous: it will lead to puzzlement about the speaker’s choice of "Must p" rather than p, but not to total communicative breakdown. This prediction seems to be in line with intuition, as well as with the experimental results reported above, which suggest that subjects find "Must p" without an argument to be substantially degraded, but not completely unacceptable.

Let me close this section by highlighting an empirical prediction of the present proposal: any construction which has the features which played an essential role in the derivation is predicted to give rise to a Support-like constraint. That is, any construction which has p as a structurally simpler

---

\(^{46}\)Importantly, "Can't p" will have the same basic update effect as a structurally simpler alternative, namely "Not p", and so we will rightly predict that "Can't p" requires a salient argument for "Not p".
relevant alternative; which has the same basic update effect as an assertion of \( p \); and which highlights the speakers’ collective doxastic relation to \( p \), is, ceteris paribus, predicted to be degraded without a salient argument for its complement (and thus also to give rise to Indirectness, via the pragmatic reasoning discussed in the last section). Further research should examine other expressions that share these features to see if this prediction is borne out.\(^{47,48}\)

### 3.5 Weakness

This concludes my main line of response to Karttunen’s Problem. Before concluding, I take up two further questions. First, I address how my account explains the felt weakness of some ‘must’-claims, and the related observation that in some cases ‘must’ is obligatory.

My account entails that ‘must’ is pragmatically strong, in the sense that an assertion of ‘Must \( p \)’ is a proposal to update the common ground with \( p \). But this account nonetheless provides a way of cashing out the intuition that ‘must’ is, in many cases, felt to weaken a claim, an intuition which has been reported in various forms in many places in the literature—though rarely with a very precise characterization of what this intuition amounts to. One notable, and, I think, telling observation made in Lassiter (2016) is that sometimes agents are happy to assert sentences along the lines of ‘Must \( p \), but I don’t know \( p \) for sure’. Lassiter produces a number of naturalistic examples like (58):

\[(58) \quad \text{This is not non-stick. There must be others but I don’t know for sure.}\]

While sentences like this sound somewhat marked to me, they are certainly preferable to the corresponding non-modal variants:

\(^{47}\) Thus for instance ‘Our evidence suggests that \( p \)’, ‘Let’s agree that \( p \)’, and ‘It is clear that \( p \)’ all seem to have these properties, and do indeed appear to be governed by corollaries of Support and Indirectness. This suggests that, even if the pragmatic derivation I have given here is mistaken in some of its details, the explanation for both Indirectness and Support is indeed pragmatic, and stems from the properties I have pointed to. See again Barker 2009; thanks to Kai von Fintel for discussion.

\(^{48}\) See Degen et al. 2015 for a different manner-implicature based approach to explaining the behavior of ‘must’ which, like the present approach, relies on the assumption that ‘Must \( p \)’ is a costly alternative to \( p \) whose use must be somehow explained. In contrast with the present approach, that approach attempts to derive Indirectness directly, rather than by way of Support, and without adverting to the different QUDs raised by modal vs. non-modal variants.
This is not non-stick. There are others but I don’t know [that] for sure.

In a closely related discussion, Degen et al. (2016) use a series of experiments to show that subjects’ production of "Must p" is associated with weaker evidence for p than their production of p alone, and likewise that they tend to infer that speakers of "Must p" have weaker evidence than speakers of p alone.

How can we make sense of this—particularly in light of the data which suggest that 'must' is pragmatically strong, an assumption which played a crucial role in our account? In uttering "Must p", on my account, a speaker proposes that her interlocutors accept p, but she proposes that they do so on the basis of an argument that is publicly available, not on the basis of her own epistemic standing. Ordinarily when a speaker asserts p, she needs to have a strong epistemic standing with respect to p: she needs to know p (or bear a similar attitude along those lines towards p), and her assertion is a proposal for her interlocutors to accept p on the basis of her knowledge—i.e. on the basis of reasons which she may not share, but which her interlocutors will generally trust she has. But asking one’s interlocutors to accept p on the basis of a shared argument, rather than on the basis of one’s own epistemic authority, will plausibly be permissible when one’s own epistemic standing towards p is something short of what would be required for an outright assertion of p alone. This is because, when the speaker proposes accepting p on the basis of a shared argument rather than her own epistemic authority, the whole group is provided the evidence in question, and tasked with determining whether or not the evidence warrants the conclusion of p—rather than simply being tasked with judging whether to trust the speaker, or not. In short, the decision whether the speaker’s evidence warrants accepting p is outsourced to the whole group.

Taking this tack may well be appropriate when the speaker thinks that her evidence warrants accepting p, but is not entirely sure, and so wants the whole group to decide whether or not p is warranted on the basis of that evidence. By sharing her evidence for p with the group, she allows the group to decide whether that evidence warrants accepting p based on that evidence—bringing closer scrutiny, from the whole group of deliberators, to the decision. Such a move will generally be warranted only when she thinks the decision merits such scrutiny; in some cases (though not all), this
will be because the evidence in question provides her with something short of certain knowledge of \( p \). Importantly, moreover, taking this tack allows the group to keep track of the inferential relations between different propositions they commonly accept. This is crucial when they think there is a chance that they will discover that one of those propositions is false, and thus will have to revise their beliefs about it—belief revision is most efficient when you have kept track of the inferential relations between your beliefs.\(^{49}\)

By contrast, when there is little chance of revision, keeping track of inferential relations just clutters the mind. And so, by contrast to "Must \( p \)", an assertion of \( p \) alone will be warranted only if the speaker is confident enough in \( p \) that she is comfortable having her interlocutors accept \( p \) based on her epistemic authority alone, i.e. only if she is comfortable putting her epistemic authority on the line in support of \( p \); and only when she thinks there is no need for her interlocutors to keep track of the inferential process that led to accepting \( p \), i.e. only when she is confident that her interlocutors will not need to remove \( p \) from the common ground.\(^{50}\)

Because of the distinctive kind of update that "Must \( p \)" amounts to, on my account—a proposal to accept \( p \) based on a shared argument—asserting "Must \( p \)" thus will be compatible with having weaker evidence for \( p \) than is required for asserting \( p \) alone, explaining the persistent intuition in the literature that there is something weak about 'must', borne out in data like Lassiter (2016)'s and Degen et al. (2016)'s.

This perspective on the felt weakness of 'must' also helps explain closely related cases, in which 'must' is felt to be obligatory. These cases, observed in Ninan 2014, also help make the reasoning spelled out so far more concrete. Consider (60):

\[
(60) \quad \text{[Suppose A and B are friends with a couple, Carl and Diane, who have been dating for a long time and are likely to get married at some point in the future. Suppose that, prior to the}
\]

\(^{49}\)In other words, when you have a sophisticated system of epistemic entrenchment. See Hansson 2016 for an overview of relevant literature.

\(^{50}\)This latter point will be especially salient for explaining the fact that "Must \( p \)" can be felt to be weak even when it is thought, or spoken in a monologue. In that case, it does not make much sense to distinguish between accepting something on the basis of the speaker's epistemic authority or not, since the speaker and the listener are the same person. But it will still matter a great deal whether you plan to accept \( p \) in a way which keeps track of the reasons for which you've accepted \( p \), or whether you are happy to simply accept \( p \) without keeping track of the inferences which led you to \( p \); opting for the former route will often be appropriate when you are less sure of \( p \), or when your evidence, and the inferential paths you have taken from it to \( p \), bear further scrutiny.
following dialogue, A and B have not heard any recent news concerning Carl and Diane’s relationship. Now consider:]

a. [A:] Carl proposed to Diane yesterday!

b. [B:] At last! She must have said ‘yes’.

c. [B:] ??At last! She said ‘yes’.  [adapted from Ninan 2014]

(60-c) is marked in comparison with (60-b). Intuitively, this is because ‘She must have said ‘yes’” makes clear that B’s conclusion is drawn from the evidence in common, whereas ‘She said ‘yes’” makes it unclear on what basis B drew her conclusion; B’s evidence is weak enough that she has every reason to think that—although she and her interlocutors now have reason to accept that Diane said ‘yes’—they may have reason, later on, to revise this commitment, and so it is crucial that they keep track of the inferential path that led to this acceptance.

This contrast is exactly what is predicted by our account. On our account, using the ‘must’ variant makes B’s inferential path explicit: by using the ‘must’ variant, B is proposing to accept that Diane said ‘yes’ based on the evidence the speakers have in common about the strength of Carl and Diane’s relationship. By contrast, using the non-modal variant leaves it open that the speaker has evidence that Diane said ‘yes’ which goes beyond what is shared in the context, thereby obscuring B’s inferential path. This, in turn, could lead to bad epistemological consequences. Suppose, for instance, that A and B discover that Carl and Diane have recently been fighting a lot. Should A revise her belief that Diane said ‘yes’? Intuitively, she should. And if B uses the ‘must’ variant, then A will know that she should, since she will know the basis of B’s conclusion that Diane said ‘yes’—she will know that B was just making an inferential leap from their shared evidence. But not so if B uses the non-modal variant without further elaboration, in which case A will not know on what basis B concluded that Diane said ‘yes’, and thus will not know whether revision is merited. (Things will be otherwise, of course, if B shares her reasoning; and, as predicted, in that case the non-modal claim is substantially improved, as in: ‘At last! She said ‘yes’—at least, that’s my conclusion based on how good their relationship is; I can’t imagine her turning him down.’)

In short: in this case, given our account of ‘must’, (60-b) amounts to a proposal to accept that
Diane said 'yes' on the basis of the evidence which is already shared; whereas (60-c) amounts to a proposal to accept that Diane said 'yes' on the basis of the speaker's epistemic authority alone. In this case, taking the latter option is, from an epistemological point of view, a dispreferred option, since it obscures the inferential path which lead to B's conclusion—an inferential path which may well be crucial for A to have access to, given the defeasibility of B’s inference that Diane said 'yes'.

The approach to 'must' which I have spelled out in this paper thus allows us to make sense of the felt weakness—and, on occasion, obligatoriness—of 'must', while simultaneously making sense of the felt pragmatic strength of 'must'. The situation is subtle: on the one hand, 'Must p' seems to have essentially the same update effect as p alone. On the other hand, assertions of 'Must p' are sometimes felt to be correlated with contexts in which the speaker’s epistemic relation to p is in some sense weaker than what is required for an assertion of p alone. Our approach predicts both these features of 'must'. On our approach, an update with 'must' is pragmatically strong: once you accept 'Must p', you also accept p. On the other hand, since, on our approach, an update with 'Must p' is a proposal to accept p on the basis of a shared argument, such an update will be most appropriate when the speaker's evidence for p follows an inferential route she wishes to make manifest to her interlocutors. That, in turn, will often be the case when her confidence in p is somewhat less than what might be required for an outright assertion of p. This accounts for the feeling that speakers can sometimes assert 'Must p' when their evidence for p is weaker than what would be required for an assertion of p alone; and it accounts for the intuition that a 'must' is sometimes felt to be obligatory, or at least preferred, when the speaker’s path to her conclusion is indirect enough that she is felt to be under some obligation to share that reasoning with her interlocutors.

---

51This explanation may extend to another class of cases, brought to my attention by Agnes Callard, involving moral judgments. 'Lee was late, and therefore disrespectful' is felicitous, but 'Lee was late, and therefore must be disrespectful' is very strange. The explanation, again, may be that we expect speakers to put the full weight of their epistemic authority behind assertions like this.
3.6 Embeddings

The final question I address concerns how to extend the present solution to Karttunen’s Problem to embeddings of ‘must’. First, note that Indirectness persists when ‘must’-claims are embedded under connectives as in (61-a) and (61-b), which communicate that the speaker’s evidence that Mark came to the party is suitably indirect:

(61) [Who came to the party?]
    a. John did, and Mark must have as well; they’re inseparable.
    b. John did, and Mark must have as well.
    c. John did and Mark did.

Second, Support also seems to be in play in cases like this. It is satisfied in (61-a), which is felicitous as it stands. (61-b) is felicitous only if we can accommodate an argument that Mark came (for instance, that John and Mark always go out together). By contrast, (61-c) requires no such inference: there need be no salient argument that Mark came to the party.

A natural first response to cases like this is that, since these signals persist through embeddings, they must be semantically encoded. But this would be too quick. On the one hand, it is not at all clear how to capture these data semantically. As I discussed in §3.3.1, encoding Indirectness as a presupposition runs into serious difficulty. And, as I discussed in §3.4.1, a lexical or presuppositional approach to Support also faces serious challenges. Encoding either of them as conventional implicatures would not do any better, since it would predict unconditional projection, and thus face the same counterexamples that make trouble for the presuppositional approaches discussed above.52 And it is clear that they are not part of the main content of ‘must’, since e.g. ‘Not must p’ clearly does not mean the same thing as ‘My evidence for p is direct, or fails to entail p’. So it is not at all obvious how a semantic approach to these signals could work.

On the other hand, we can explain these cases pragmatically by taking essentially the same

---

52This is so at least on the approach due to Potts 2005; others have however since resisted the claim that conventional implicatures project unconditionally (e.g. Martin 2016), which may create room for maneuver here.
strategy as in the unembedded cases. The details of this will depend on an account of how ‘must’ embeds; I will not try to give an explicit account of this here, but I will sketch the basic idea. The basic idea is that ‘must’ embedded under connectives quantifies over the worlds in ‘must’s local context. Thus, for instance, the ‘must’ in (61-b) quantifies over the worlds in the global context where John came to the party. (61-b) will thus have essentially the same update effect as (61-c), and so the former will be in competition with the latter, which is structurally simpler. We can thus reason, as in the unembedded case, that the speaker is choosing to highlight the agents’ collective doxastic relation to the prejacent, and thus that there is an argument commonly available for the prejacent (at least once we take into account the information in its local context—in this case, that John came). It is widely thought that Non-Redundancy holds not only at the level of sentences, but also at a local, sub-sentential level. We can then reason as above to show that the relevant argument must be indirect: the argument must be shared (by the time we reach the local context), and thus the prejacent should not follow from it in an overly direct way, at risk of being locally redundant. This approach extends straightforwardly to other embeddings of epistemic modals under connectives, provided there exists a non-modal variant which communicates the same basic information.

The indirectness of embeddings of ‘must’ under attitudes, like (62), can be explained in a simple but slightly different way.

(62) James thinks Bob must be in his office.

Here, again, the ‘must’ seems to suggest that James’ evidence is indirect, though no correlate of Support seems to apply here. How to predict this will depend on the interaction between attitude verbs and ‘must’. The most natural way to treat epistemic modals under attitudes (as in Stephenson 2007a,b) is to take the agent of the attitude predicate to “anchor” the modal base, so that ‘must’ is interpreted as quantifying over the set of worlds compatible with a relevant body of the agent’s evidence. Thus an attitude ascription like (62) says that James believes his evidence entails that Bob

53On local contexts, see e.g. Stalnaker 1970, Karttunen 1974, Heim 1982.
54Such a theory is required to explain the infelicity of sentences like ‘If John is in Paris, then John is in Paris and he’s having fun’. See e.g. Schlenker 2009, Katzir and Singh 2013, Mayr and Romoli 2016.
55Importantly, this approach will not predict that ‘must’ under negation requires an argument, since “Not must p” and “Not p” do not have the same basic update effect.
is in his office. A comment on what kind of evidence an agent thinks she has will, in turn, only generally be worth making if that agent has some beliefs about her evidence. But, in general, an agent will only bother to reflect on her evidence if her evidence is in some sense indirect—if there is some epistemic space between her evidence and the conclusion in question.

More needs to be said to spell out these accounts in detail—in particular, a thorough account of embedded epistemic modals must be spelled out. But this discussion should suffice to give a sense of how the present account makes sense of the persistence of Indirectness (and, to a degree, Support) signals for embedded ‘must’.

3.7 Conclusion

The argument of this chapter came in three stages. I began by arguing that, in addition to Indirectness, we need Support to fully characterize the differences in felicity conditions between an assertion of ‘Must p’ and an assertion of p alone (as well as the difference between ‘Must p’ and epistemic modal claims of other strengths). Next, I argued that we can account for Indirectness by deriving it pragmatically from Support. Finally, I argued that we should account for Support as a manner implicature.

The three main parts of this argument are, to a degree, independent: the second and third part depend on the first, but not vice versa; and the second and third part are independent of each other. If each of these moves is successful, however, then taken together, they constitute a solution to Karttunen’s Problem: characterizing and explaining the differences in felicity conditions between an assertion of ‘Must p’ and an assertion of p. In short: because of its more complex form and its difference in subject matter from p alone, an assertion of ‘Must p’, unlike an assertion of p alone, requires that an argument be given for p. From this requirement, in turn, we can conclude that the speaker’s evidence for p is indirect (in a relevant sense), on the basis of general considerations about the flow of information in conversation. I have argued that this approach provides an explanatorily and empirically satisfying account of Support; of the subtle pattern of judgments underlying Indirectness; and, finally, of the sense in which ‘must’ is felt to be weak, in spite of its pragmatic
strength.

In conclusion, I highlight three broad upshots of my approach to Karttunen's Problem. The first is about the relation between 'must' and evidentiality. I have argued that 'must' does not grammaticalize a certain constraint on the type of evidence which the speaker must have for its prejacent. Rather, the felt indirectness of 'must'-claims is accounted for pragmatically; and it is accounted for not directly in terms of judgments about intuitive categories of evidence, but rather in terms of judgments about *redundancy*, explaining the subtle and surprising patterns of judgments about when *Indirectness* is satisfied.

The second is about the meaning of 'must'. My derivation of *Support* rests on a certain (independently motivated) assumption about how we generally interpret 'must'-claims. If the derivation is successful, then it provides further motivation for adopting that assumption.

The third regards the theory of redundancy. I have argued that the norm that governs redundancy in assertions is stronger in some respects and weaker in others from the norm that has been assumed in the literature. A theory of redundancy will play a central role in understanding how our minds structure and access information, and judgments about discourses provide a rich source of data for these theories.

I close with an abstract point about the architecture of semantic and pragmatic theories. My proposal rests on the assumption that an assertion of "Must p" and an assertion of p have the same basic update effect, but different semantic values. Indeed, on the semantics I have sketched, "Must p" and p informationally entail one another—in the sense that a context updated with either one entails the other—but they do not semantically entail one another. The possibility of this divergence between update effect and meaning proved essential for capturing the subtle divergence in felicity conditions between "Must p" and p alone; and so, if we are to adopt an approach like this, then—contrary to some lines of thought in dynamic semantics—we must distinguish semantic content from pragmatic update effect in our theorizing about natural language.
Chapter 4

Modality and expressibility

In this concluding chapter, I take an abstract perspective on the choice of semantic theories for epistemic modals. I develop formal tools to compare the expressive power of different theories of epistemic modals with respect to the embedding data which they can account for. I show that the relational theory is strictly more expressive than several recent revisionary theories, including the domain, update, and state-based semantics. This necessitates a substantial shift in perspective on the debate regarding the semantics of epistemic modals. On the one hand, since essentially any data that one of these revisionary theories can make sense of can also be handled by the standard theory, this shows that it is very unlikely that there are data which will show the standard theory to be wrong and a revisionary theory correct. On the other hand, insofar as those revisionary theories are weaker than the standard theory vis-à-vis their expressive power, there should be a prejudice in their favor, as long as they can make sense of all the relevant data. In the final part of the chapter, I explore the choice between the domain framework and the standard framework in light of these results, focusing on the behavior of modals under attitudes.

4.1 Introduction

In recent years, philosophers and linguists have discovered data which shed new light on the embedding behavior of epistemic modals—data of the sort discussed in the first chapter. Those data
have in most cases been used to argue against the standard theory of epistemic modals—on which epistemic modals have, roughly, the semantics of modal operators in standard modal logics—and in favor of various revisionary theories.

In this chapter I take an abstract perspective on the underlying dialectic in this debate. Whenever embedding data are used to argue against semantic theory $A$ and in favor of semantic theory $B$, it is important to ask whether $A$ could, after all, make sense of those data. It is possible to ask that question on a case-by-case basis. But suppose we could show that $A$ can make sense of all the embedding data which $B$ can make sense of. This would, in one fell swoop, undermine all arguments in favor of $B$ on the basis of embedding data. And, provided that the converse does not hold—that is, that $A$ can make sense of strictly more embedding data than $B$ can—it would also provide a new form of argument in favor of $B$: this kind of result would show that $B$ is, in a precise sense, more constrained than $A$, suggesting that there are pro tanto simplicity-based considerations in favor of $B$.

In this chapter I develop formal tools which allow us to make comparisons of this kind, which I call comparisons of relative potential expressibility. I motivate the development of these tools through discussion of the recent debate about epistemic modals, and show their use by applying them to that debate. These tools show that several prominent revisionary theories which have been developed in response to embedding data are strictly less expressive than the standard "relational" theory, in the sense, again, that the relational theory can make sense of any embedding data involving epistemic modals which those theories can make sense of, but not vice versa (within very mild limits). This necessitates a fundamental reorientation in how to think about the choice between these frameworks. Indeed, the situation is roughly the opposite of the way it is standardly presented. On the one hand, we need not worry about discovering data which the relational theory cannot make sense of, but which these other theories can: we can rest assured that the relational theory can account for all the data which those theories can account for. And, conversely, these facts tell us that we may well discover data that the relational theory can make sense of, but that the revisionary theories cannot. On the other hand, the fact that these revisionary theories are expressively weaker than the relational theory in the sense in question suggests that there should be some ceteris paribus
prejudice in their favor: if natural language makes use only of a strict subset of the expressive power of the relational semantics, this is an important fact which should be reflected in our semantic theory.

Besides advancing the specific debate I am concerned with, concerning epistemic modals, this discussion demonstrates the importance of exploring facts about relative potential expressibility in comparing different semantic models of fragments of natural language; my discussion only begins to explore the underlying mathematical formalism, and possible applications of that formalism.

My plan is as follows. I begin, in §4.2, by discussing data from Yalcin 2007 which Yalcin took to merit a revisionary account of epistemic modals, the domain semantics, and then reviewing a recent response in Ninan 2016 which shows that Yalcin’s revisionary account of these data can be exactly replicated in the standard relational framework. In §4.3, I develop the tools of relative potential expressibility to show that this result is no accident: it follows from much more general facts about the relative potential expressive power of the domain versus relational semantics. I extend this result with a variety of further comparisons, including of the update semantics, state-based semantics, and the extensions of the relational, domain, and update semantics to quantified languages. In every case, I show that relational framework is the most expressive option. In §4.4, I explore the significance of these results. In §§4.5–4.6, I make this discussion more concrete by exploring the choice between the domain and relational frameworks in light of these results, focusing on the behavior of modals under attitudes and emphasizing the importance of a kind of methodological parsimony in the choice between semantic frameworks. I conclude in §4.7.

4.2 Background

I will start by reviewing the standard relational semantics, Yalcin (2007)’s data, his response to those data, and finally a recent response to Yalcin in Ninan 2016.

4.2.1 The relational theory

On the standard relational theory, epistemic modals denote quantifiers over a set of worlds: those worlds which are accessible from the world of evaluation. (For simplicity, throughout this chapter I
will focus on the standard relational semantics presented, rather than on the local relational semantics given in Chapter 1; at a relevant level of abstraction, there is no difference in expressive power between these, and so for present purposes we can ignore differences between them.) We can think of the accessible worlds as determined by a binary accessibility relation between worlds—or, equivalently and somewhat more conveniently, a modal base: a function from worlds to sets of worlds. The modal base for epistemic modals is meant in some sense to track the contextually relevant information (though in what sense precisely won’t matter for us). ‘Might’ is treated as an existential quantifier, so ‘Might p’ is a claim that the contextually relevant information leaves it open that p;1 ‘must’ is treated as the dual universal quantifier, so ‘Must p’ is a claim that the contextually relevant information entails p. More formally:2

**Definition 4.2.1. Relational Semantics:** For any modal base f, possible world w, and sentence p:

- \([\text{Might } p]^{f,w} = 1 \iff \exists w' \in f(w) : \lbrack p \rbrack^{f,w'} = 1\)

- \([\text{Must } p]^{f,w} = 1 \iff \forall w' \in f(w) : \lbrack p \rbrack^{f,w'} = 1\)

1 Except when working with a logical language below, I will use roman letters to stand for sentences of natural language and italic letters to stand for the corresponding propositions (sets of possible worlds), implicitly relativized to a modal base (in this framework) or information state (in the domain framework). Where p is a proposition, \(\lnot p\) is its complement. I will leave off a context parameter throughout just for readability; insofar as the languages we are working with contain context-sensitive terms (beyond epistemic modals), these should be read as implicit throughout.

2 This view is due in particular to Kratzer 1977, 1981, building on earlier work in modal logic, e.g. Kripke 1963 (though see Footnote 14 for important divergences in the present discussion from the standard Kripkean approach). This presentation simplifies Kratzer's theory in two ways, one irrelevant (for Kratzer, a modal base is a function from worlds to sets of propositions, not worlds), and one possibly relevant (for Kratzer, modals are evaluated relative to a second parameter, an ordering source). Insofar as an ordering source plays in interesting role in what follows, though, it makes the relational theory even more expressive than it already is, and thus adding an ordering source throughout will not change the expressibility hierarchies that we will construct in what follows (ordering sources are most relevant for dealing with epistemic modals of intermediate strength, like ‘likely’ and ‘probably’, which bring up interesting further issues which I do not have space to discuss here). Another close variation on this implementation has modal bases supplied by a variable assignment to a variable associated with the modal claim (see von Fintel 1994); adopting that variation would, again, make no difference to our results. Note that these semantic entries are commitments about the semantic values of epistemic modal claims, not about their assertoric value; the present discussion remains neutral on the debate between contextualist, expressivist, and relativist interpretations of these entries, which are debates about assertoric value, not semantic value (On the distinction see e.g. Lewis 1980; Ninan 2010; Rabern 2012, 2013; for relativist approaches, see Egan et al. 2005; Stephenson 2007b; MacFarlane 2011, 2014; for expressivism, Yalcin (2007); Swanson (2015); Moss (2015) and others). In including modal base parameters in the index of evaluation, I am assuming that they are shiftable by embedding operators; this assumption is standard in treatments of the relational semantics (as in e.g. Kratzer 1991).
4.2.2 Yalcin (2007)'s challenge

Yalcin (2007) observed that, when you embed a sentence with the form \( \text{flp} \) and might not \( \text{p}^{-1} \) (a right-modal-nested Wittgenstein sentence, which he calls an *epistemic contradiction*) under 'Suppose', the result is infelicitous.\(^3\)

(1)  
   a. #Suppose it's raining and it might not be.  
   b. #Suppose (p and might not p).

As we discussed briefly in §1.7.1, the issue for the standard approach is that it seems to predict that a sentence like (1-a) will be perfectly felicitous. To spell out the problem, let's focus (as Yalcin does) on the corresponding declarative sentence. A sentence like (2) is felt to attribute inconsistent suppositions to Alfred, a fact which explains why the imperative variant will be felt to be infelicitous (since there is something wrong with commanding someone to make their suppositions inconsistent):

(2)  
   a. Alfred supposes it's raining and it might not be.  
   b. A supposes (p and might not p).

But, given the relational semantics, it seems like (2-a) will just say that Alfred supposes that it's raining, and that some contextually salient body of evidence is compatible with the proposition that it's not raining. In other words, (2-a) should be roughly equivalent to (3), which does not attribute any kind of inconsistency to Alfred.

(3) Alfred supposes that it's raining and that, for all he knows, it's not raining.

To make this worry more precise, let's spell out the assumptions about connectives and attitude predicates in the background. First, Yalcin assumes that connectives have standard, Boolean semantics.\(^4\)

\(^3\)Yalcin notes that a similar phenomenon arises for epistemic contradictions in the antecedent of conditionals; we will not discuss that further data point here, but, as the generalization below shows, insofar as the domain framework can account for those data, the relational framework can as well.

\(^4\)Although Yalcin does not emphasize this assumption, it is important to note, since his puzzle does not arise if we make different assumptions—for instance, assuming dynamic conjunction.
Definition 4.2.2. Boolean connectives, relational version:

- \([p \text{ and } q]_{f,w} = 1 \iff [p]_{f,w} = 1 \text{ and } [q]_{f,w} = 1\)
- \([\text{Not } p]_{f,w} = 1 \iff [p]_{f,w} = 0\)

Second, Yalcin assumes a Hintikkan approach to attitude predicates (Hintikka, 1962). On this approach, attitude predicates denote universal quantifiers over the possible worlds compatible with the relevant attitude of the agent in question. That is, for any attitude verb \(V\), where \(V_{A,w}\) denotes the worlds compatible with everything that \(A\)'s:

Definition 4.2.3. Hintikka semantics, relational version:

- \([A \text{ V's } p]_{f,w} = 1 \iff \forall w' \in V_{A,w} : [p]_{f,w'} = 1\)

Informally: "A V's p\(^\top\) is true at a world \(w\) and modal base \(f\) just in case \(p\) is true at every world compatible with what \(A\)'s, relative to \(f\). Thus, in particular, "A supposes \(p\)" just says that \(p\) is true at every world compatible with A's suppositions, relative to the relevant modal base. Together with the relational semantics for epistemic modals and Boolean semantics for connectives, we thus predict that "A supposes (p and might not p)\(^\uparrow\)" as evaluated at modal base \(f\) and world \(w\), says that every world in A's supposition state makes \(p\) true (relative to \(f\)), and that every world in A's supposition state can access under \(f\) some world where \(p\) is false (relative to \(f\)). So (2-a) should just mean something like (3)—that Alfred supposes that it's raining, and he supposes that it is compatible with some contextually salient body of evidence that it's not raining—and thus should not be felt to ascribe incompatible suppositions to Alfred.

4.2.3 Yalcin (2007)'s domain semantics

This is a puzzle for the combination of the relational semantics with the Hintikka semantics for attitude verbs and Boolean semantics for connectives. Yalcin (2007) responded to this puzzle by developing a theory that is revisionary twice over: it rejects the relational theory of epistemic modals, as well as the Hintikka semantics for attitude verbs. Yalcin, following an early version of MacFarlane 2011, adopts a domain semantics for epistemic modals. On the domain semantics, epistemic
modals denote quantifiers over a set of worlds (or information state) \( s \) which is supplied, not by an accessibility relation, but rather as a world-independent parameter of the index.

**Definition 4.2.4.** *Domain semantics:*

- \([\text{Might } p]^{s,w}=1 \iff \exists w' \in s : [p]^{s,w'} = 1\)
- \([\text{Must } p]^{s,w}=1 \iff \forall w' \in s : [p]^{s,w'} = 1\)

In short: "Might \( p \)" says that \( p \) is compatible with a given information state; "Must \( p \)" says that \( p \) is entailed by a given information state. This semantics is very close to the relational semantics, but with a set of worlds substituted where the relational semantics has a function from worlds to sets of worlds.

Yalcin then generalizes the standard Boolean semantics for the connectives in the obvious way:

**Definition 4.2.5.** *Boolean connectives, domain version:*

- \([p \text{ and } q]^{s,w}=1 \iff [p]^{s,w} = 1 \text{ and } [q]^{s,w} = 1\)
- \([\neg p]^{s,w}=1 \iff [p]^{s,w} = 0\)

Yalcin, finally, develops a new semantics for attitude verbs, which builds on the Hintikka semantics, but stipulates that the attitude verb supplies the set of attitude worlds as the domain of quantification for its complement. For any attitude verb \( V \), we have:

**Definition 4.2.6.** *Yalcin semantics for attitude verbs:*

- \([A V's \ p]^{s,w}=1 \iff \forall w' \in V_{A,w} : [p]^{V_{A,w},w'} = 1\).

This combination of views solves the puzzle sketched just now. Consider a sentence with the form of (4):

\[(4) \quad \text{A supposes (p and might not p).}\]

\(^5\)Yalcin presents a semantics only for 'supposes', but suggests that it should be extended to other attitudes in the obvious way; for simplicity, I am presenting the general format here. Likewise for Ninan's semantics below.
Where $S_{A,w}$ is the set of worlds compatible with $A$'s suppositions at $w$, on Yalcin’s approach to attitude verbs, (4) will be true just in case, at every world $w'$ in $S_{A,w}$, $p$ and might not $p'$ is true relative to $\langle S_{A,w}, w' \rangle$. Suppose that a sentence with the form of (4) is true. Then, given this approach to attitudes, together with the domain semantics and the Boolean semantics for the connectives, we know that:

- for every world $w'$ in $S_{A,w}$, $p$ is true at $\langle S_{A,w}, w' \rangle$; and
- for every world $w'$ in $S_{A,w}$, there is some world $w''$ in $S_{A,w}$ such that $p$ is false at $\langle S_{A,w}, w'' \rangle$.

Clearly these two conditions can only both be met if $S_{A,w}$ is empty, rendering the universal quantification here vacuous. Thus a sentence with the form of (4) will ascribe inconsistent suppositions to $A$, and thus the imperative version will be infelicitous because it is a command to make an inconsistent supposition.\(^6\)

4.2.4 Ninan (2016)’s response

Yalcin’s proposal is, again, revisionary in two respects: it couples a new semantics for modals (the domain semantics) with a new semantics for attitude verbs. Ninan (2016) shows that there is a way to exactly replicate the predictions of Yalcin’s framework regarding the interaction between attitude verbs and epistemic modals, while maintaining the relational semantics for epistemic modals. For any set of worlds $s$, let $f^s$ be the constant function which takes any world to $s$. Then, for attitude verb $V$ and modal base $f$, Ninan proposes:

**Definition 4.2.7.** Ninan semantics for attitude verbs:

- $[A V's p]\_{f,w} = 1$ iff $\forall w' \in V_{A,w} : [p]\_{f^{V_{A,w}}, w'} = 1$.

\(^6\)Although the semantics I have presented here is Yalcin (2007)’s, the explanation I have given diverges somewhat from Yalcin’s own explanation. Yalcin claims that, on his semantics, a sentence with the form “$A$ supposes ($p$ and might not $p$)” does not ascribe contradictory suppositions to $A$, but is rather itself a contradiction (i.e. true at no point $\langle s, w \rangle$). But this does not follow from his semantics, at least as it stands: “$A$ supposes ($p$ and might not $p$)” will be true at $\langle s, w \rangle$ whenever $A$’s suppositions at $w$ are inconsistent. We could, however, get to this point by adding to the semantics of attitude verbs a presupposition that the attitude is consistent, as Ninan (2016) proposes (this is true for any of the semantics I consider below, as well). I am not convinced that we need to do this, rather than rely on pragmatic reasoning, but that question is orthogonal to my main interests here.
To see how this account makes sense of Yalcin’s data, consider its predictions about a sentence with the form “A supposes (p and might not p)”\(^7\). Given Ninan’s semantics for attitude verbs, together with the relational semantics for modals and classical entries for the connectives, a sentence with this form will be true at a world \(w\) and modal base \(f\) just in case:

(i) for every world \(w'\) in \(S_{A,w}\), \(p\) is true at \(\langle f^{S_{A,w}}, w' \rangle\); and

(ii) for every world \(w'\) in \(S_{A,w}\), there is some world \(w''\) in \(f^{S_{A,w}}(w')\) such that \(p\) is false at \(\langle f^{S_{A,w}}, w'' \rangle\).

But since \(f^{S_{A,w}}\) takes every world to \(S_{A,w}\), (ii) amounts to the following:

(iii) for every world \(w'\) in \(S_{A,w}\), there is some world \(w''\) in \(S_{A,w}\) such that \(p\) is false at \(\langle f^{S_{A,w}}, w'' \rangle\).

These conditions are clearly never all met when there are any worlds compatible with A’s suppositions; they will only be met when \(S_{A,w}\) is empty, rendering the quantification in these conditions vacuous. Thus “A supposes (p and might not p)”\(^7\) will be felt to be equivalent to the claim that A’s suppositions are inconsistent, furnishing an explanation for the infelicity of the imperative version.\(^7\)

4.3 Relative potential expressibility

Is it a happy accident that we can reproduce Yalcin’s system in a relational framework? In other words, is it a fluke of Yalcin’s semantics for attitude verbs which enabled us to reproduce this interaction in a relational framework? This is an important question for both practical and theoretical reasons. Practically, this is an important question because, as we saw in the first chapter, Yalcin’s data are not the only challenge to the relational semantics from embedding behavior. Yalcin 2007 discusses similar data involving epistemic contradictions in the antecedents of conditionals, and a long line of research on epistemic modals in quantified environments has raised similar challenges for relational approaches.\(^8\) The result of Ninan’s discussed just now shows that the relational semantics can make sense of just one fragment of these data. One possibility from a practical point of view

\(^7\)Cf. Schultz 2010; Holliday and Icard 2017 for related results which show that the logic of the domain semantics is a restricted case of ordinary modal logic.

would be to go through each data point, and see whether we can replicate the revisionary accounts in each case in a relational framework.\footnote{Ninan (2016) takes one further step in that direction, showing that a similar move to the one summarized above can account for epistemic contradictions in the antecedents of conditionals in a relational framework; the discussion in the first chapter is, of course, another move in this direction.} But this approach is inefficient; and, more problematically, even if we could show that all the known data can be accounted for in the relational framework, this would still leave open the possibility that more data are lurking undiscovered which some revisionary approach can account for, but which the relational semantics cannot account for. It would be far more consequential if we could show that the relational framework can account for any embedding data which can be accounted for in the domain framework (or other revisionary approaches). From a more theoretical point of view, this is an important question because it tells us something about the structure of the underlying relationship between the relational framework and its competitors.

In this section I will spell out a framework which allows us to answer questions like this, by allowing us to compare the expressive power of two theories of a given fragment with respect to what embedding operators are definable (in a compositional manner) within those two theories.

Ordinarily it is straightforward to compare two semantic theories for a given language $L$. Provided that the intensions in the two theories are the same kind of thing, we can simply compare the intensions that those theories assign to the sentences of $L$. If the theories agree everywhere, then there is nothing to choose between them (at least empirically; more abstract considerations may, of course, still bear on the choice); otherwise, we will find some sentence $\varphi$ of $L$ and index $i$ such that one theory will say that $\varphi$ is true at $i$, while the other theory says that $\varphi$ is false at $i$. We will then choose between the theories by trying to determine which prediction is correct. This is by no means always easy to do, but it is at least clear in broad outlines how such a comparison is meant to go.

But it is harder to compare semantic theories of a language $L$ when the kind of thing that an intension is differs across two theories. And this is just the situation for the domain and relational semantics. An index (the object relative to which sentences are true or false) in the relational framework is a pair of a modal base and a world; an index for the domain semantics is a pair of an information state and a world. This makes it difficult to directly compare the two frameworks. We can still, however, compare theories like this with respect to what sentence operators are definable
in each theory. In particular, suppose that we add an arbitrary new \( n \)-place sentence operator to the language, and extend the first semantic theory to give a compositional semantics for the resulting language (a semantics which assigns to the new operator as its semantic value a function from \( n \)-tuples of intensions to a new intension). Suppose that we will be guaranteed to be able to find a way of extending the second semantic theory to give a compositional semantics for that operator such that the operator has exactly the same logic, according to the second theory, as it does according to the first theory. Then we can say that the first theory is no more expressive than the second theory with respect to embedding data: any embedding data which the first theory can make sense of can also be accounted for in the second theory.

We can spell this out more precisely as follows:

**Definition 4.3.1. Relative Potential Expressibility:** Given two models \( \mathcal{M} \) and \( \mathcal{M}' \) for a language \( \mathcal{L} \), \( \mathcal{M} \) is no more expressive than \( \mathcal{M}' \) with respect to \( \mathcal{L} \) (written \( \mathcal{M} \leq_{\mathcal{L}} \mathcal{M}' \)) iff, for any set of new sentence operators \( \mathcal{O} \), for any extension \( \mathcal{M}^{\mathcal{O}} \) of \( \mathcal{M} \) to \( \mathcal{L}^{\mathcal{O}} \), there is an extension \( \mathcal{M}'^{\mathcal{O}} \) of \( \mathcal{M}' \) to \( \mathcal{L}^{\mathcal{O}} \) which preserves the logic of \( \mathcal{O} \) from \( \mathcal{M}^{\mathcal{O}} \).

By ‘model’ I mean something fairly rich (somewhat richer than what is usually intended in model theory): a domain of worlds and atomic valuation function, plus a set of indices (which will generally be constructed from the set of worlds), plus an interpretation function which assigns truth value to all sentences relative to an index. This rich notion of a model corresponds closely to the kinds of systems that semanticists construct to make sense of a fragment of natural language, and will be the most useful notion to work with in what follows. A language is just a set of sentences. A sentence operator is any operator whose intension is a function from the intensions of the sentences it embeds to a new intension—in other words, any operator which is given a broadly compositional semantics. The extension of a language with a new set of sentence operators is the closure of the language under those operators. Finally, what it means for \( \mathcal{M}^{\mathcal{O}} \) to preserve the logic of \( \mathcal{O} \) from \( \mathcal{M}^{\mathcal{O}} \) is the following: for any \( n \)-place sentence operator \( O^n \in \mathcal{O} \), for any sentence \( \psi \) in \( \mathcal{L}^{\mathcal{O}} \) and \( n \)-tuple of sentences \( \varphi \) from \( \mathcal{L}^{\mathcal{O}} \), \( (O^n(\varphi) \models_{\mathcal{M}^{\mathcal{O}}} \psi) \leftrightarrow (O^n(\varphi) \models_{\mathcal{M}'^{\mathcal{O}}} \psi) \). For a set of sentences \( \Phi \) and sentence \( \psi \), \( \Phi \models_{\mathcal{M}} \psi \) means, in turn, that \( \psi \) is true in \( \mathcal{M} \) at every index where all the elements
of \( \Phi \) are true in \( \mathcal{M} \). All of this is spelled out more formally in Appendix C, but this should suffice for the casual reader.

The notion of relative potential expressibility spelled out here is somewhat different from extant notions of expressibility, which are typically concerned with comparisons between different languages rather than with comparisons between different models of the same language with respect to arbitrary extensions of the language. But the notion of relative potential expressibility is just the notion we need to answer the questions posed at the beginning of this section. To see this, suppose that we can show that the domain semantics is no more expressive (in the sense spelled out here) than the relational semantics (as we will do in a moment). Then we can show that, for any way of defining an attitude operator in the domain semantics, there is a way of defining an attitude operator in the relational semantics which has exactly the same logic. In light of this, Ninan’s result is just one instance of a very general phenomenon. Likewise, suppose that we can show that the relational semantics is strictly more expressive in this sense than the domain semantics (as we will do in a moment). It follows that there are operators we can define in the relational semantics which have logical properties which cannot be replicated for any operator the domain semantics.

It turns out that relative potential expressibility boils down to something very simple: the existence of a certain kind of function between the indices of the two models in question:

**Fact 4.3.1. Characterization of Expressibility:** For any models \( \mathcal{M} \) and \( \mathcal{M}' \) and language \( \mathcal{L} \), \( \mathcal{M} \preceq_{\mathcal{L}} \mathcal{M}' \) iff there is function \( g \) from the indices of \( \mathcal{M} \) to those of \( \mathcal{M}' \) which is such that (i) for any sentence \( \varphi \) of \( \mathcal{L} \) and index \( i \) in \( \mathcal{M} \), \( \varphi \) true at \( i \) in \( \mathcal{M} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}' \); and (ii) \( g \) is an injection.

The proof of Fact 4.3.1, and of all the other facts stated in the rest of this section, are not essential to

---

10 I write \( \models_{\mathcal{M}} \psi \) for \( \{ \varphi \} \models_{\mathcal{M}} \psi \).

11 A close corollary, brought to my attention by Harvey Lederman, is that, if \( \mathcal{M} \preceq_{\mathcal{L}} \mathcal{M}' \), then, for any extension \( \mathcal{L}^0 \) of \( \mathcal{L} \) and any extension of \( \mathcal{M} \) to a new model \( \mathcal{M}^0 \) of \( \mathcal{L}^0 \), there is a way of extending \( \mathcal{M}' \) to a model \( \mathcal{M}'^0 \) of \( \mathcal{L}^0 \) such that the logic of \( \mathcal{L}^0 \) in \( \mathcal{M}^0 \) is a subset of the logic of \( \mathcal{L}^0 \) in \( \mathcal{M}'^0 \) (the logic of a language \( \mathcal{L} \) in a model \( \mathcal{M} \) is the set of pairs \( (\Phi, \psi) \) such that \( \Phi \in \mathcal{L} \wedge \psi \in \mathcal{L} \) and \( \Phi \models_{\mathcal{M}} \psi \)). Construct \( \mathcal{M}'^0 \) in such a way that there is an injection \( g \) which preserves truth of sentences of \( \mathcal{L}^0 \) between \( \mathcal{M}^0 \) and \( \mathcal{M}'^0 \); we can do this by using the method sketched in the proof of Fact 4.3.1. If \( \Phi \models_{\mathcal{M}'^0} \psi \), then \( \psi \) is true at every index \( i \) in \( \mathcal{M}'^0 \) where all the elements of \( \Phi \) are, so \( \psi \) is true in \( \mathcal{M}^0 \) at every \( \mathcal{M} \)-index in the image of \( g^{-1} \) where all the elements of \( \Phi \) are (where \( g^{-1} \) is the inverse of \( g \), which will be well-defined on \( g \)'s image since \( g \) is an injection), since \( g^{-1} \) will preserve truth for the sentences of \( \mathcal{L}^0 \) between \( \mathcal{M}^0 \) and \( \mathcal{M}'^0 \) where defined. \( g^{-1} \) will be a surjection from a subset of \( \mathcal{M}' \)-indices to the \( \mathcal{M} \)-indices, and so \( \psi \) will be true in \( \mathcal{M}^0 \) everywhere that the elements of \( \Phi \) are all true in \( \mathcal{M}^0 \) and so \( \Phi \models_{\mathcal{M}^0} \psi \).
my main points, and thus are relegated to Appendix C. This characterization result greatly simplifies the proofs of the facts that follow, and, I think, provides some intuitive grip on those results: whether one model is less expressive than another\(^\text{12}\) depends on the structural relations between the models, and in particular whether the first model can, from a very abstract perspective, be embedded into the second model in a truth-preserving way.

4.3.1 The domain semantics

With this discussion in hand, let us turn our attention to the specific comparisons of expressive power which we are concerned with here: namely, comparisons between different semantics for epistemic modals. Let our language \(L^\diamond\) comprise atomic sentences \(p_i : i \in \mathbb{N}\) (which I will write \(p, q, r\)\(^\text{13}\)) together with sentences of the form \(\Diamond \varphi\), for any sentence \(\varphi \in L^\diamond\) (\(\Diamond\) abbreviates 'might').

Let \(R\) be the model of the relational semantics, and \(D\) the model of the domain semantics. Recall that an index in \(R\) is a pair \((f, w)\), with \(f\) a modal base. An atomic sentence \(p\) is true in \(R\) at \((f, w)\) just in case \(v(p, w) = 1\), and \(\Diamond \varphi\) is true in \(R\) at \((f, w)\) just in case there is a world \(w'\) in \(f(w)\) such that \(\varphi\) is true at \((f, w')\).\(^\text{14}\) And recall that an index in \(D\) is a pair \((s, w)\), with \(s\) an information state. An atomic sentence \(p\) true in \(D\) at \((s, w)\) just in case \(v(p, w) = 1\). A sentence with the form \(\Diamond \varphi\) true at \((s, w)\) in \(D\) just in case there is a world \(w'\) in \(s\) such that \(\varphi\) is true at \((s, w')\). I will assume that \(R\) and \(D\) have the same stock of worlds and valuation function.

With this in mind, we can show that the domain semantics is no more expressive than the relational semantics:

**Fact 4.3.2.** \(D \preceq_{L^\diamond} R\).

\(^\text{12}\)When I say 'less/more expressive' I always have in mind relative potential expressibility.

\(^\text{13}\)In discussion of natural language, earlier and later in the chapter, I use roman letters to stand for any sentences of natural language, and italics to stand for corresponding propositions; in this more formal section and the proofs in Appendix C, I use italic letters to stand for atomic sentences, and Greek letters to stand for arbitrary sentences of the language.

\(^\text{14}\)In a more traditional modal Kripke semantics, a modal base/accessibility relation would not be included as an element of the index, but would rather be hardwired into the model in question. Then we would talk about the class of models for the relational semantics, instead of the model of relational semantics. I am here following instead the broad strategy of Lewis 1980 in including in the index any shiftable parameter of evaluation. Treating modal bases as elements of the index in this way will greatly facilitate comparison with other systems. Even given this approach, it is of course a further simplifying assumption that there is one intended model of the various semantics I explore below, rather than a class of models constrained by the interpretive stipulations characteristic of those semantics. If we relax that assumption, then my results can be reformulated as follows: for any domain model \(D\), there is a relational model \(R\) such that \(D \preceq_{L^\diamond} R\), and so on.
The proof strategy is to construct a function $g$ which takes any index $(s, w)$ in $D$ to $(f^s, w)$, where $f^s$ is the constant function from worlds to $s$. It is easy to show that this function preserves truth and is an injection.\footnote{Note that it is easy to extend this to a proof that the domain semantics is no more expressive than the relational semantics with respect to an expanded language containing not just atoms and 'might' sentences, but also sentences formed with the Boolean connectives. Let $\mathcal{L}^{\Diamond,B}$ denote this language, and let $\mathcal{D}^B$ and $\mathcal{R}^B$ denote the domain and relational models coupled with the classical semantics for the connectives given above. Then we can show:}

Next note that the relational semantics is not less expressive than the domain semantics with respect to $\mathcal{L}^{\Diamond}$:

**Fact 4.3.4.** $\mathcal{R} \not\prec_{\mathcal{L}^{\Diamond}} \mathcal{D}$.\footnote{We write $\mathcal{M} \preceq_{\mathcal{L}} \mathcal{M}'$ as shorthand for $\mathcal{M} \preceq_{\mathcal{L}} \mathcal{M}' \land \mathcal{M}' \not\preceq_{\mathcal{L}} \mathcal{M}$.}

Fact 4.3.4 follows immediately from the observation that it is a theorem of $\mathcal{D}$ that $\Diamond \varphi \leftrightarrow \Diamond(\Diamond \varphi)$, but not a theorem of $\mathcal{R}$. This will make it impossible to find a truth-preserving injection from the indices of $\mathcal{R}$ to those of $\mathcal{D}$, since at some indices of $\mathcal{R}$, $\Diamond \varphi$ will be true while $\Diamond(\Diamond \varphi)$ is not, and vice versa. Thus from Facts 4.3.2 and 4.3.4, we have that the domain semantics is strictly less expressive than the relational semantics with respect to $\mathcal{L}^{\Diamond}$.\footnote{Though see Moss 2015 for discussion of nested epistemic modals.}

**Fact 4.3.5.** $\mathcal{D} \prec_{\mathcal{L}^{\Diamond}} \mathcal{R}$.\footnote{Though see Moss 2015 for discussion of nested epistemic modals.}

It is worth noting that 'might's are not generally able to stack in natural language: that is, sentences with the form 'Might (might p)\textsuperscript{2}' are generally infelicitous.\footnote{Note that it is easy to extend this to a proof that the domain semantics is no more expressive than the relational semantics with respect to an expanded language containing not just atoms and 'might' sentences, but also sentences formed with the Boolean connectives. Let $\mathcal{L}^{\Diamond,B}$ denote this language, and let $\mathcal{D}^B$ and $\mathcal{R}^B$ denote the domain and relational models coupled with the classical semantics for the connectives given above. Then we can show:}

Thus from Facts 4.3.2 and 4.3.4, we have that the domain semantics is strictly less expressive than the relational semantics with respect to $\mathcal{L}^{\Diamond}$.\footnote{Note that it is easy to extend this to a proof that the domain semantics is no more expressive than the relational semantics with respect to an expanded language containing not just atoms and 'might' sentences, but also sentences formed with the Boolean connectives. Let $\mathcal{L}^{\Diamond,B}$ denote this language, and let $\mathcal{D}^B$ and $\mathcal{R}^B$ denote the domain and relational models coupled with the classical semantics for the connectives given above. Then we can show:}

**Fact 4.3.6.** $\mathcal{R} \not\prec_{\mathcal{L}^{\Diamond,B}} \mathcal{D}$.\footnote{Note that it is easy to extend this to a proof that the domain semantics is no more expressive than the relational semantics with respect to an expanded language containing not just atoms and 'might' sentences, but also sentences formed with the Boolean connectives. Let $\mathcal{L}^{\Diamond,B}$ denote this language, and let $\mathcal{D}^B$ and $\mathcal{R}^B$ denote the domain and relational models coupled with the classical semantics for the connectives given above. Then we can show:}

For the sake of simplicity and generality, however, I will focus on our simpler language $\mathcal{L}^{\Diamond}$ in what follows.
We know this because we can show that any function from the indices of $\mathcal{R}$ to those of $\mathcal{D}$ which preserves truth for sentences of $\mathcal{L}^{\#}$ will fail to be an injection. Consider two relational indices with the same world and different modal bases. Now suppose those modal bases have the same value at the world of evaluation. Any such function will in some cases have to map both such indices to the same domain index, and thus will fail to be an injection; then Fact 4.3.7 follows by Fact 4.3.1.

By contrast, we will obviously still have that $\mathcal{D}$ is no more expressive than $\mathcal{R}$ with respect to $\mathcal{L}^{\#}$, and so we have:

**Fact 4.3.7.** $\mathcal{D} \prec_{\mathcal{L}^{\#}} \mathcal{R}$.

### 4.3.2 The update and state-based semantics

In the next section, I will explore the upshots of these results. An important fact about the formal tools developed here, however, is that it is straightforward to apply them to further comparisons: nothing in the kind of comparison we are doing here depends on details of the domain or relational semantics. In the rest of this section, I will illustrate this by exploring the relative potential expressive power of a few more important semantics of epistemic modals which have been motivated by various embedding data. In each case, the relational semantics comes out as the most expressive option.

Let me note that the results I summarize here depend on assumptions about the background sets of worlds and valuation functions. Those assumptions are, I think, reasonable; readers should consult Appendix C for details.

First, the comparison of expressive power can be extended to the update semantics. In the update semantics $\mathcal{U}$, due to Veltman (1996), building on Heim 1982, 1983, the intension of a complex sentence is a dynamic object: a function from information states (usually called contexts in this literature) to information states. The most natural way to think about this approach from the point of view of the formal framework we are using is to think of our "indices" as pairs of contexts, with $\varphi$ "true" at a pair $\langle s, c \rangle$ in $\mathcal{U}$ just in case $\langle s, c \rangle \in [\varphi]_\mathcal{U}$; we continue to think of entailment as preservation of truth, so that $\varphi \models_U \psi$ just in case the function denoted by $\varphi$ in $\mathcal{U}$ is a subset of a function denoted by $\psi$ in $\mathcal{U}$ (likewise, *ceteris paribus*, for multi-premise entailment). Entailment
in the dynamic framework is more typically spelled out in terms of \textit{acceptance} (as in Veltman 1996), where a context \( c \) accepts a sentence \( \varphi \) just in case \( \varphi \) is a fixed point for \( \varphi \), i.e. just in case \( [\varphi]_U(c) = c \). Then we can say that \( \varphi \) entails \( \psi \) in the dynamic sense, just in case whenever \( \varphi \) is accepted by any context \( c \), so is \( \psi \). Our notion of entailment is stronger than this one: whenever we have \( \varphi \models_U \psi \), then we have that \( \varphi \) entails \( \psi \) in the dynamic sense, but not \textit{vice versa}. Not everything in the update framework, however, is definable in terms of acceptance—for instance, the update treatment of ‘might’ is not.\(^{18}\) So I believe the stronger approach to entailment—which treats pairs of contexts, rather than single contexts, as our “indices”—is the appropriate way to think about the update semantics from the point of view of our formalism. (For discussion of a system where dynamic acceptance plays the role of “truth”, see the discussion of the state-based semantics below.)

With this background, we can say that an atomic sentence \( p \) is true\(_U \) at \((s, c)\) just in case the result of removing all worlds \( w \) from \( s \) such that \( v(p, w) = 0 \), where \( v \) is our valuation function (which we can assume, again, is the same as for the update and domain frameworks). A sentence of the form \( \Diamond \varphi \) “tests” whether \( \varphi \) is compatible with the input context: \( \Diamond \varphi \) is true\(_U \) at \((s, c)\) just in case, roughly, \( s \) includes a \( \varphi \)-world and \( s = c \), or \( s \) doesn’t include a \( \varphi \)-world, and \( c = \emptyset \) (more precisely, just in case \( [\varphi]_U(s) \neq \emptyset \) and \( s = c \), or else \( [\varphi]_U(s) = \emptyset \) and \( c = \emptyset \)). The update semantics is no more expressive than the relational semantics with respect to \( \mathcal{L}^\Diamond \):

\textbf{Fact 4.3.8.} \( U \preceq_{\mathcal{L}^\Diamond} R \).

The proof goes by way of constructing an injection from pairs of contexts to relational indices which preserves truth for all sentences in \( \mathcal{L}^\Diamond \). The injection goes by way of distinguishing a variety of different cases, and is not particularly intuitive; in my view, Fact 4.3.8 is quite a surprising fact.

We can also show that the relational semantics is more expressive than the update semantics:

\textbf{Fact 4.3.9.} \( R \not\preceq_{\mathcal{L}^\Diamond} U \).

The proof is identical to the proof of Fact 4.3.4 (that \( R \not\preceq_{\mathcal{L}^\Diamond} D \)): it is a theorem of \( U \) that \( \Diamond p \leftrightarrow \Diamond (\Diamond p) \), but not a theorem of \( R \). Thus we have:

\(^{18}\)To see this, note that in standard update systems, \( \Diamond (\Diamond \varphi \land \neg \varphi) \) is sometimes accepted and sometimes not accepted at non-empty contexts, even though \( \Diamond \varphi \land \neg \varphi \) \textit{never} is accepted at non-empty contexts. So the semantics of \( \Diamond \) cannot be given in terms of acceptance.
Fact 4.3.10. $U \prec_{L^0} R$.

Interestingly, though, we do not have parallel results for the domain semantics:

Fact 4.3.11. $U \not \prec_{L^0} D$.

We can prove this by showing that there is no truth-preserving injection from the indices of $U$ to those of $D$. Furthermore, we can use the same method to show that the domain semantics is not less expressive than the update semantics:19

Fact 4.3.12. $D \not \prec_{L^0} U$.

Thus the domain and update semantics are expressively incommensurable—though both are strictly less expressive than the relational semantics.

Note, finally, that these relations are all preserved for $L^{0-}$, the language which does not allow stacked 'might's. First, we have $D \not \prec_{L^{0-}} U$ (the proof is the same as the proof of Fact 4.3.12). That, together with Fact 4.3.7 (that $D \prec_{L^{0-}} R$), and the fact that $\preceq$ is transitive (see Fact 4.3.15 below), shows that $R \not \prec_{L^{0-}} U$; otherwise, we would have $D \preceq_{L^{0-}} U$. Finally, the proof that $U \preceq_{L^0} R$ extends immediately to $L^{0-}$, so we have:

Fact 4.3.13. $U \prec_{L^{0-}} R$.

The last system I will discuss here is the state-based semantics of Hawke and Steinert-Threlkeld 2016.20 The state-based semantics $S$ is very similar to the update semantics, except an index is a single set of worlds. The state-based semantics is thus much simpler than the update semantics, though it retains many of the attractions of the update semantics. This simplicity, however, impacts the expressive power of the state-based semantics. In the state-based semantics, an atomic sentence $p$ is true at a set of worlds $c$ in $S$ just in case, for every world $w \in c : v(p, w) = 1$, where $v$, again, is the valuation function of the state-based semantics. And a sentence of the form $\lozenge \varphi$ is true in $S$

---

19These facts are particularly interesting in relation to the limited equivalence between the update and domain semantics proved in Rothschild 2017. See Rothschild and Yalcin 2015, 2016 for detailed discussion of dynamic semantics and their relation to static semantics.

20It is a more complicated matter to extend the comparison to the bilateral state-based semantics in Aloni 2016, Steinert-Threlkeld 2017, Chapter 3.
just in case \( \varphi \) is true in \( S \) in some \( \{w\} \subseteq c \). The state-based semantics, unlike its close cousin the update semantics, is strictly less expressive than the domain semantics:

**Fact 4.3.14.** \( S \prec_{L_0} D \).

Furthermore, for any language \( L \), \( \preceq_L \) is transitive; this follows from Fact 4.3.15:

**Fact 4.3.15.** For any language \( L \), \( \preceq_L \) is a partial pre-order.

So we also have:

**Fact 4.3.16.** \( S \prec_{L_0} R \).

### 4.3.3 Quantified modal languages

The final comparison of expressive power which I will explore here concerns quantified modal languages. This is a particularly interesting topic for present purposes because epistemic modals embed in fascinating ways in the scope of quantifiers—ways that, as we discussed above, have been used to argue against the standard semantics and in favor of various quantified enrichments of the update semantics. The present expressibility results can, however, be extended to show that, if we extend all these frameworks to quantificational languages in the most straightforward way, the expressive hierarchies for the non-quantified language remain unchanged.

To show this, consider a standard quantificational language \( L_0^\Diamond \), built out of a vocabulary comprising variables \( x_i : i \in \mathbb{N} \); \( n \)-place relation symbols \( R^n_i : i \in \mathbb{I} \) for every \( n \geq 0 \); and a one-place sentence operator \( \Diamond \). \( L_0^\Diamond \) is the smallest set comprising atomic sentences with the form of an \( n \)-place relation symbol \( R^n \) followed by an \( n \)-tuple of variables; and sentences of the form \( \Diamond \varphi \), for any \( \varphi \in L_0^\Diamond \). We can extend the basic relational model \( \mathcal{R} \) to a model \( \mathcal{R}_0^\Diamond \) for this quantified language by adding a domain to the model, and adding a variable assignment (a function from variables to elements of the domain) to our indices, so that they amount to triples comprising a variable assignment, accessibility relations, and world. Our valuation function \( v \) now takes a world and an \( n \)-place relation symbol and returns an \( n \)-place relation on the domain. Our truth clauses for atomic sentences and \( \Diamond \) will be generalized in the usual way: where \( \langle a, f, w \rangle \) is a relational index, with \( a \) a variable
assignment, $f$ a modal base, and $w$ a possible world, we say that an atomic sentence of the form $R^n((x_1, x_2, \ldots, x_n))$ is true$_R$ at $(a, f, w)$ iff $(a(x_1), a(x_2), \ldots, a(x_n)) \in v(R^n, w)$. Our semantics for $\Diamond$ remains essentially unchanged: $\Diamond \varphi$ is true$_R$ at $(a, f, w)$ iff there is a world $w' \in f(w)$ such that $\varphi$ is true$_R$ at $(a, f, w')$. We can treat quantifiers as sentence operators (shifting variable assignments, in the usual way) which we can freely add to our language. We can likewise enrich the domain semantics to a model $D^3$ of $L^3$ in a parallel fashion, augmenting our indices so they are triples of variable assignments, information states, and worlds, and extending the interpretation function in the obvious way. It is clear that these changes do not affect the expressive hierarchy between the domain and relational semantics. That is, we have:

**Fact 4.3.17.** $D^3 \prec_{L^3} R^3$.

The proof is a straightforward generalization of the proof of the parallel result for the non-quantified case.

Things get more interesting when we turn to the update semantics. The most obvious way to incorporate quantification into the update framework is to simply treat intensions as functions from a variable assignment to a function from contexts to contexts (as in Yalcın 2015), so that $U^3$-indices have the form $(a, (s, c))$, for $a$ a variable assignment and $s$ and $c$ contexts. With $v$ our valuation function, generalized as for the relational and domain semantics, we say that, for atomic sentence $p$ of the form $R^n((x_1, x_2, \ldots, x_n))$, $p$ is true$_R$ at $(a, (s, c))$ iff $c$ is the result of removing from $s$ all and only worlds $w$ such that $(a(x_1), a(x_2), \ldots, a(x_n)) \notin v(R^n, w)$. And we say that $\Diamond \varphi$ is true$_R$ at $(a, (s, c))$ iff $[\varphi]_R(a)(s) \neq \emptyset \land s = c$, or $[\varphi]_R(a)(s) = \emptyset = c$. If we go this way, then, once again, the expressive hierarchy from above will be preserved: the quantified update semantics will be strictly less expressive than the quantified relational semantics.

**Fact 4.3.18.** $U^3 \prec_{L^3} R^3$.

Interestingly, this is not the most common way to incorporate quantification into the dynamic framework in which the update semantics is cast. In the standard quantified extension of that framework, developed in Heim 1982, 1983, indices are treated, not as pairs of a variable assignment and...

---

21I'll set aside the state-based semantics for now; I don't know of quantificational versions of that semantics.
a pair of information states, but rather as pairs of 'files', where a file is a set of world-variable assignment pairs. These complications may well change the expressive hierarchies; I will not explore these facts further here. But it is clear that these complications have to do with the richer way in which those systems introduce quantification into the update setting as compared with the way quantification is standardly introduced into the relational or domain setting, not with the update semantics for epistemic modals. Moreover, I suspect that it may be possible to recapture the expressive hierarchy from above by taking an equally rich approach to quantification in the relational framework (in other words: we may well be able to combine the relational approach to epistemic modals with the characteristic dynamic approach to quantification), though I will not explore that here.

4.4 The upshot

There is much more to explore regarding the relative expressive power of different semantics for epistemic modals, but I will stop here. Let me briefly summarize my findings, and then emphasize a few upshots of this discussion.

I have shown that the (quantified) domain semantics, (quantified) update semantics, and state-based semantics are strictly less expressive than the relational semantics with respect to a language comprising 'might'-claims and atomic sentences. This shows that, for any way of extending our language with a new set of sentence operators and any way of extending the (quantified) domain, (quantified) update, or state-based models to give semantics for those operators, we can extend the relational framework to these operators in a way which exactly replicates the logic of that operator. But the converse is not true: there are operators definable in the relational framework whose logic cannot be replicated in the domain, update, or state-based frameworks.

This discussion thus shows that, there are no embedding data involving 'might' which the relational semantics cannot make sense of, but which the domain, update, or state-based semantics can make sense of, within mild bounds: as long as we focus on sentences where 'might' takes as a complement just sentences in our simple starting language. In fact, of course, 'might' can embed
sentences of much more complexity. But, as far as the embedding behavior of epistemic modals goes, this limitation seems to be harmless; all of the data that I know of which have been used to motivate departures from the relational framework stay within these bounds. Moreover, many of the results above can be easily extended to more complex starting languages.

This shows that, in the context of these mild limitations, there are no data involving embedded epistemic modals that the relational semantics for epistemic modals cannot make sense of, but which the domain or update semantics can. And that, in turn, necessitates a substantial change in perspective on the dialectic in the debate about the correct semantics for epistemic modals. At a high level, that dialectic has been presented roughly as follows. First, a conservative assumption in favor of the relational semantics is generally taken for granted. Then, this assumption is challenged by data which the relational semantics apparently has trouble accounting for. Finally, the data are used to advocate a revisionary semantics which can make sense of them.

The present results show that this gets things exactly backwards. On the one hand, these results show that there will not be embedding data which tell against the relational framework and in favor of the domain, update, or state-based framework. But there may well be data which tell the other way: embedding data which the relational framework can make sense of, but which the domain, update, and state-based frameworks cannot make sense of.

On the other hand, these results show that there should be no prejudice in favor of the relational semantics. If natural language does not make use of the full expressive power of the relational semantics, we should record this fact in our semantic theory: this would provide a pro tanto reason to opt for less expressive frameworks. The less expressive a theory of ‘might’ is, the fewer stipulations are needed to make it match the data from natural language. Thus, if the less expressive frameworks can make sense of the behavior of embedded epistemic modals, then the present expressibility results show that there should be a simplicity-based prejudice in their favor. (That prejudice should, of course, only be pro tanto; there may well be further considerations which still tell in favor of a more expressive framework. For instance, a more expressive framework may be able to account for the data in a more natural way than a less expressive framework.22)

22Cf. the situation in syntax, where transformational grammars were used for many decades even though all the data
In short: it is not that the relational theory should be presumed to be correct, and must fight a defensive action to try to make sense of surprising new data. On the contrary, there should be a prejudice in favor of weaker revisionary theories—but these theories must show that they are sufficiently powerful to make sense of the embedding behavior of epistemic modals.

4.5 Simple shifty systems

This concludes my discussion of expressibility. My plan for the rest of this chapter is to briefly compare two of the frameworks under discussion—the domain and relational semantics—with these expressibility results in mind. My goal here is not to settle the choice between these frameworks, but rather to point to the kinds of considerations that should bear on choosing between two semantic frameworks like these—in particular to advocate a certain kind of methodological parsimony in choosing between semantic theories. I should note that the discussion which follows is, however, strictly independent from everything that has come so far.

Let’s begin by considering the system given in Ninan (2016). It is easy to see that Ninan’s system—his semantics for attitude predicates, plus the relational semantics—is equivalent to Yalcin’s system—his semantics for attitude predicates, plus the relational semantics—in the sense that, for any world \( w \), information state \( s \), and modal base \( f \), \( \diamond V p \) is true in Yalcin’s system at an index \( (s, w) \) just in case it is true in Ninan’s system at the index \( (f, w) \). Ninan’s account thus shows that we can not only make sense of Yalcin’s data in a relational framework, but that we can in fact exactly replicate the system that results from the domain semantics, plus Yalcin’s semantics for attitude reports, in the relational framework.

This means that there is nothing to distinguish these two views in terms of empirical coverage of the meaning of modals under attitudes. So, if we are to choose one of these frameworks, then the expressibility results in the last section suggest that Yalcin’s system is to be preferred. (This is so available at the time could be captured in a context-free grammar, because of the ease with which transformational grammars captured the relevant data. Thanks to Roni Katzir for discussion and for pointing out this connection.

---

2 This follows from the respective semantics for attitude verbs, plus an induction on the length of formulas which shows that, in the context of an ordinary language containing the Boolean connectives (given their classical semantics) and ‘might’, for any \( p \) \( p \) is true in the domain semantics at \( \langle s, w \rangle \) just in case it is true in the relational semantics at \( \langle f^*, w \rangle \) (assuming the same stock of worlds and atomic valuations in both systems).
at least if we focus only on the behavior of epistemic modals under attitudes; considerations about epistemic modals in other environments may of course yield a different perspective, but I will focus here on modals under attitudes, for the sake of having a manageable empirical domain.)

But this is not the end of the story. Yalcin’s system and Ninan’s systems have serious empirical shortcomings. In the rest of this section I will bring out three problems for any system which make the same predictions about modals under attitudes as Yalcin’s and Ninan’s systems (call such systems simple shifty systems).24

The first, and most serious, observed in Yalcin 2012b; Dorr and Hawthorne 2013, concerns modals under factive attitude verbs. According to the simple shifty account, for non-modal p, ‘A knows might p’ will be true just in case p is compatible with A’s knowledge. Thanks to the factivity of knowledge, any truth will be compatible with A’s knowledge. And so for any non-modal p and agent A, if p is true, then p will be compatible with A’s knowledge, and thus ‘A knows might p’ is predicted to be true. In other words, for non-modal p, the inference from p to ‘A knows might p’ is valid in the simple shifty framework.

This is obviously wrong: it is possible to fail to know that some truth might obtain. People often have false beliefs. Suppose that John sees Mark enter his office and close the door. John believes Mark is in his office. Indeed, he has every reason to think that Mark is in his office and to think that he knows this with total confidence. If we asked John where Mark is, he would say: ‘Mark’s in his office.’ If we asked him if he was sure, he would say, ‘Yes. I know that Mark is in his office. I’m absolutely certain about it.’ But now suppose that, unbeknownst to John, Mark has a secret exit in the floor of his office, and has used this exit to leave the office and go to the bar. In this situation, (5) seems false, and (6) true:

\begin{align*}
(5) & \text{ John knows that Mark might be at the bar. } \\
(6) & \text{ John doesn’t know that Mark might be at the bar. }
\end{align*}

But, since Mark is in fact at the bar, the simple shifty approach wrongly predicts that (5) is true, and

\footnote{In addition to Yalcin’s and Ninan’s systems, we could get to a simple shifty framework in a variety of other ways: for instance, via the update semantics, when combined with the approach to attitudes in Heim 1992, or the event-relative modal and attitude semantics of Hacquard 2006, 2010 (see also Anand and Hacquard 2013).}
(6) is false.\textsuperscript{25}

The prediction of the simple shifty account provides sufficient reason to reject it. I’ll point out two more problems here, however, mainly to provide desiderata for positive accounts. First, this framework gets the entailment relations between $\lnot A$ knows might $p$ and $A$ believes might $p$ wrong. In this framework, the first says that $p$ is compatible with $A$’s knowledge; the latter says that $p$ is compatible with $A$’s beliefs (if $A$’s beliefs are consistent). Whenever $p$ is compatible with someone’s beliefs, it is compatible with their knowledge; but not vice versa. So, as Dorr and Hawthorne (2013) observe, in this framework, $A$ believes might $p$ and $A$’s beliefs are consistent\textsuperscript{26} entails $A$ knows might $p$.\textsuperscript{26} But $A$ knows might $p$ can be true while $A$ believes might $p$ is false.

Both these predictions look wrong. First, one can consistently believe that something might be the case, without knowing that it might be the case; many birthers consistently believe that Obama might have been born in Kenya, but they certainly don’t know this. Second, we are generally happy to infer from someone knowing that something might be the case to their believing it might be the case; to see this, note the oddness of conjunctions with the form $A$ knows might $p$, but $A$ doesn’t believe might $p$, as in (7).\textsuperscript{27}

\begin{equation}
\text{(7) } \text{John knows that Mark might be in his office, but he doesn’t believe that Mark might be in his office.}
\end{equation}

The final issue I’ll point out is that the simple shifty system predicts that ‘must’ is vacuous under attitudes.\textsuperscript{28} That is, sentences with the form $A$ V’s must $p$ and $A$ V’s $p$ are predicted to

\textsuperscript{25}In response to this problem, Yalcin (2012b) proposes a revision to the simple shifty framework; cf. Beaver 1992, 2001; Rothschild 2011; Willer 2013; Moss 2016 for relevantly similar approaches. Those approaches avoid this problem, but face the third problem I sketch below.

\textsuperscript{26}I focus here on non-modal $p$, but this is just for simplicity; these points hold also for modal $p$. Dorr and Hawthorne (2013) report this entailment without the caveat that $A$’s beliefs must be consistent, but if $A$’s beliefs are inconsistent, the inference won’t go through: if $A$’s beliefs are inconsistent, then $A$ believes might $p$ will be trivially true, even if $p$ is not compatible with their knowledge.

\textsuperscript{27}This is not entirely uncontroversial; see Hawthorne et al. (2016); Bledin and Lando (2017); Beddor and Goldstein (2017). The cases discussed there require careful consideration, which I do not have space for here. Nothing essential turns on the assumption that ‘knows might’ entails ‘believes might’, though.

\textsuperscript{28}See Hacquard 2010, §6.1.2 for an observation of a parallel issue for unembedded modals in her system, attributed to Paul Portner.
be semantically equivalent—to express the same proposition at any context. This is because in the simple shifty semantics, the only contribution of the embedded modal is quantificational force—but ‘must’ and attitude predicates have the same quantificational force, namely universal force. These predicted equivalences, however, are wrong. A construction with the form "A knows/believes must p" is generally only felicitous if A's evidence for p is in some sense indirect.29 For instance—to modify a stock example—suppose that Sue is watching it rain, and on this basis concludes that it's raining out. We can describe the situation with (8), but not with (9):

(8) Sue knows/believes it's raining.

(9) Sue knows/believes it must be raining.

By contrast, suppose that Sue can’t see outside, but sees some of her colleagues come inside with wet umbrellas, and on this basis concludes that it’s raining out. We can then describe the situation with either (8) or (9). The simple shifty semantics cannot explain this difference, since, again, on that approach, (8) and (9) are semantically equivalent.

Defenders of the simple shifty approach may want to explain these data by means of a conservative modification of the semantics of modality they are assuming, for instance by adopting a presupposition of indirectness along the lines suggested by von Fintel and Gillies (2010). But there are a few problems with this approach. First, it is not at all clear how to marry von Fintel and Gillies (2010)'s approach with the domain semantics, or Ninan's constant-function version of the relational semantics. Second, as Ippolito (2017) and Chapter 3 discuss, indirectness does not project like a presupposition—so it is not clear, even if we could manage this, that we should. Nor will we have better luck encoding this indirectness signal as a conventional implicature or part of the main content of modal claims, for reasons von Fintel and Gillies (2010) discuss. So it does not look to me like there is a conservative modification available within a broadly simple shifty framework to avoid the present issue.

29 There is not much discussion on the topic of ‘must’ under attitudes, but the present hypothesis generalizes a common parallel observation about unembedded ‘must’; see Karttunen (1972), Veltman (1985), Kratzer (1991), von Fintel and Gillies (2010), Kratzer (2012b), Matthewson (2015), Lassiter (2016), Giannakidou and Mari (2016), Sherman (2016), Mandelkern (2017), and Chapter 3 of this dissertation. For specific discussion of embedded ‘must’ see Ippolito (2017).
4.6 Weaker constraints

In light of all this, the simple shifty system—whether implemented in the domain or relational frameworks (or, for that matter, the update or event-relative frameworks)—looks quite unattractive. Given Yalcin’s data, this seems to put us in a bind. Can we make sense of those data, without making grossly implausible predictions elsewhere in our system?

One thing that Ninan (2016)’s system makes clear is that replicating the simple shifty system in the relational framework requires a very strong constraint on modal bases for modals under attitudes: in Ninan’s approach, modal bases are constant functions to the worlds compatible with the attitude. This raises a natural question: could we account for Yalcin’s data with a weaker constraint, one that avoids the implausible predictions of the simple shifty framework? I will approach this question by investigating the weakest constraint which would suffice to predict Yalcin’s data. We can look for such a constraint within both the domain and relational frameworks. This will give us a sense of what constraints those data compel us to adopt within each framework, and, therefore, to what degree we are compelled to adopt the simple shifty approach if we are to make sense of Yalcin’s data in one of those frameworks.

In answering this question, I will assume some background constraints. First, I will continue to assume that our connectives are the standard Boolean ones, as above. While adopting non-standard, ‘dynamic’ connectives would on its own explain Yalcin’s data, it would not extend to the order and scope variants like the following, which are all just as bad as Yalcin’s sentences:

(10)  a. #Suppose it might be raining and it isn’t.
     b. #Suppose (might p and not p).

(11)  a. #Suppose it’s raining and suppose it might not be.

---

30A slightly weaker constraint than Ninan’s would suffice—namely, that modals under attitudes take any world compatible with the attitude to the set of worlds compatible with the attitude. But this is still a very strong constraint.

31This is so even if we adopt an enriched dynamic system like that in Yalcin 2015; Rothschild and Klinedinst 2015, or a symmetric dynamic conjunction like that given in the first chapter; while, as I argued there, a symmetric dynamic conjunction is independently motivated by other data, it does not on its own help with these data. Having said that, adopting such a conjunction does change the space of possibilities for the kinds of constraints that account for Yalcin’s data, as I discuss briefly in Footnote 32; for simplicity, I will ignore those complexities here, as my main goal at present is methodological.
b. \#Suppose (p and might not p).

(12) a. \#Suppose it might be raining and suppose it isn’t.
b. \#Suppose (might p and not p).

And so for our purposes we can ignore complexity in the meaning of connectives, and assume simple Boolean entries.

Second, I will continue to assume that attitudes are represented as sets of possible worlds, as in the Hintikkan framework. It is interesting to explore what happens when we relax this assumption, but I will set aside such approaches, partly for simplicity, and partly because it is relatively clear what philosophical sense we can make of broadly Hintikkan approaches, and much less clear what sense we could make of revisionary approaches (e.g. approaches which treat an attitude state as a set of pairs of a modal base and world, or sets of information states).

4.6.1 Weaker constraints in the relational framework

Given these background conditions, let us ask first what the weakest constraint is which is required to make sense of Yalcin’s data—that is, to ensure that ‘A supposes (p and might not p)’ is equivalent to ‘A supposes (p and not p)—in the relational framework. The constraint turns out to be the following: provided A’s suppositions are consistent, then some supposition world can only access other supposition worlds. In other words, where \(f\) is the accessibility relation relative to which we evaluate modals under ‘A supposes’, our constraint runs as follows: \(S_{A,w} \neq \emptyset \rightarrow \exists w' \in S_{A,w} : f(w') \subseteq S_{A,w}\). Call this the subset relational constraint. We can schematically encode this constraint as follows:

**Definition 4.6.1. Subset relational semantics:**

- \([A \forall V’s p]^{f,w}\)
  - a. defined only if \(V_{A,w} \neq \emptyset \rightarrow \exists w' \in V_{A,w} : f(w') \subseteq V_{A,w} \quad \text{subset relational constraint}
  - b. where defined, true iff \(\forall w' \in V_{A,w} : [p]^{f,w'} = 1 \quad \text{standard Hintikkan truth conditions}\)
This is not the only way of encoding the subset relational constraint; a more elegant approach would be roughly in the style of the bounded theory spelled out in the first chapter (which entails a very similar constraint, in a more principled way). But the present implementation is very simple, and it will suffice for the purposes of this chapter. Call the semantic system that results from this entry, plus the relational semantics for modals, the subset relational system.

Here’s why the subset relational constraint is sufficient to account for Yalcin’s data. Suppose that \( \neg A \supposes (p \text{ and might not } p) \) is true at \( \langle f, w \rangle \). Then (i) every supposition world \( w' \) is such that \( p \) is true at \( \langle f, w' \rangle \); and (ii) every supposition world \( w' \) is such that there is some world \( w'' \in f(w') \) such that \( p \) is false at \( \langle f, w'' \rangle \). Suppose further that A’s suppositions are consistent. Then we know that (iii) some world in \( S_{A,w} \) can only access worlds in \( S_{A,w} \) under \( f \). But then (ii) tells us that some supposition world \( w' \) is such that \( p \) is false at \( \langle f, w' \rangle \). That contradicts (i). And so A’s suppositions must be inconsistent, contrary to our assumption. Thus, as desired, we predict that \( \neg A \supposes (p \text{ and might not } p) \) entails that A’s suppositions are inconsistent. (It is straightforward to confirm that this explanation extends to order and scope variations, as well as to other attitude predicates.)

Here’s why the subset relational constraint is necessary to account for Yalcin’s data. If the constraint is violated, then we will always be able to find a value for \( p \) such that \( \neg A \supposes (p \text{ and might not } p) \) is true, even though A’s suppositions are consistent. Suppose the subset relational constraint is violated for some agent A, world \( w \), and modal base \( f \): that is, \( S_{A,w} \) is non-empty, and \( \forall w' \in S_{A,w} : f(w') \setminus S_{A,w} \neq \emptyset \). Let \( p \) denote \( S_{A,w} \). By construction, every world in \( S_{A,w} \) is a \( p \)-world, and every world in \( S_{A,w} \) will be able to access a \( \neg p \)-world under \( f \). Thus \( \neg A \supposes (p \text{ and might not } p) \) is true at \( w \), even though A’s suppositions at \( w \) are consistent.

Thus, given our background assumptions, the subset relational constraint is the weakest constraint which accounts for Yalcin’s data in the relational framework. The constraint implicit in Ninan’s simple shifty semantics—that modal bases must be constant functions to the set of attitude world—turns out to be much stronger than necessary to account for the data. And thus the relational semantics is by no means locked into a simple shifty approach if it is to make sense of

---

32 The constraint encoded there is closely related to, but logically orthogonal, the subset relational constraint; the reason we were able to find a logically orthogonal constraint which also accounts for Yalcin’s data is because we gave conjunction a non-classical semantics.
Yalcin's data—a fortunate fact, given the problems for the simple shifty approach.

Furthermore, strikingly, the subset relational approach avoids all three problems sketched above for the simply shifty approach. First, on the subset relational approach, non-modal $p$ does not entail $\gamma A$ knows might $p^-$—by contrast to the simple shifty semantics, where, again, this is a valid inference. To see this, suppose $p$ is true at some world $w$. The factivity of knowledge entails that $w$ is compatible with A's knowledge at $w$, for any A. The subset relational constraint entails that some world in $K_{A,w}$ can access only other worlds in $K_{A,w}$. But it does not follow that every world in $K_{A,w}$ can access $w$ (or indeed that any world can), and thus it does not follow that $\gamma A$ knows might $p^-$ is true at $w$.

Second, in the subset relational system, knowledge entails belief, but not vice versa—capturing the inferential relations standardly assumed between knowledge and belief, unlike the simple shifty approach. If $\gamma A$ knows might $p^-$ is true relative at $(f, w)$, then every world compatible with A's knowledge in $w$ can access a $p$-world under $f$; and so, given the standard assumption that every world compatible with an agent's beliefs is compatible with her knowledge, it follows that every world compatible with A's beliefs in $w$ can access a $p$-world under $f$, and thus that $\gamma A$ believes might $p^-$ is true at $(f, w)$. But the converse entailment does not hold, since all of A's belief worlds can access a $p$-world without all of A's knowledge worlds being able to access a $p$-world.

Finally, the subset relational system predicts that $\gamma A V's must p^-$ and $\gamma A V's p^-$ are logically orthogonal: either can be true while the other is false. The first says that every one of A's V-worlds can access only $p$-worlds; the latter says that every one of A's V-worlds are $p$-worlds. These conditions are entirely orthogonal. This is a substantial improvement over the simple shifty predictions that $\gamma A V's must p^-$ is semantically equivalent to $\gamma A V's p^-$. There is, of course, more to be done to marshal these facts into an explanation of how 'must' patterns under attitude verbs. For instance,

---

33 Here's a countermodel. Suppose that $K_{A,w} = \{w, w'\}$ with $p$ is true at $w$ and false at $w'$; let $f(w) = \{w\}$ and $f(w') = \{w'\}$. Then $f$ satisfies the subset relational constraint; $p$ is true at $w$; but $\gamma A$ knows might $p^-$ is not true at $w$, since some world in A's knowledge state, namely $w'$, cannot access a $p$-world.

34 Again, I focus on non-modal $p$, but this holds in full generality.

35 A countermodel: suppose that $K_{A,w}$ again is $\{w, w'\}$, and $B_{A,w} = \{w\}$, with $f$ again taking each world to the singleton set containing it and $p$ true at $w$, false at $w'$. Then $\gamma A$ believes might $p^-$ is true at $(f, w)$ while $\gamma A$ knows might $p^-$ is false at $(f, w)$.

36 Suppose that $V_{A,w} = \{w, w'\}$, with $f(w) = \{w\}$ and $f(w') = \{w''\}$. Suppose $p$ is true at $w$ and $w''$, and false at $w'$. Then $\gamma A V's must p^-$ is true at $(f, w)$, while $\gamma A V's p^-$ is false. Conversely, suppose that $p$ is true at $w$ and $w'$ and false at $w''$. Then $\gamma A V's p^-$ is true at $(f, w)$, while $\gamma A V's must p^-$ is false.
we will want to say how modals under attitudes are generally interpreted in order to make sense of
the fact that \( \forall A V's \) must \( p \) is generally only felicitous when A’s evidence is in some sense indirect
(presumably their interpretation tracks something about the agent’s evidence, in some broad sense,
which would go some way towards such an explanation). And we will want to explore whether there
are independent constraints on accessibility relations, and perhaps on the relation between an indi-
vidual’s beliefs and what they believe about their evidence, which might predict that \( \forall A V's \) must
\( p \) will in many cases be true only when \( \forall A V's \) \( p \) is, or vice versa. I will not undertake further exploration of these complex issues here; what is important for present purposes is that the present approach is capable of providing the right kind of foundation for a story like this—in sharp contrast to the simple shifty approach, which, again, predicts that ‘must’ makes no semantic contribution under attitudes.

4.6.2 Weaker constraints in the domain framework

Let’s return now to the domain semantics. It is natural to ask whether—just as Ninan replicated
Yalcin’s domain system in the relational framework—we can replicate the subset relational system
in the domain framework. The answer is negative. Given our background assumptions (Boolean
connectives, plus representing attitudes as sets of worlds), any semantics for attitudes in the domain
framework will have the property that \( \forall A V's \) might (\( p \) or \( q \)) entails \( \forall (A V's \) might \( p \)) or \( \forall (A V's \) might \( q \)).\(^{37}\) But this inference pattern will be invalid in the subset relational semantics.\(^{38}\) Thus (given our background assumptions) any domain semantics for attitude verbs will validate an inference pattern which the subset relational system does not, and so there is no way to replicate the latter system in the domain framework. This is, of course, unsurprising from the point of view of the expressibility results that we reviewed earlier, which showed that the relational framework

\(^{37}\)Proof: Consider a point \( \langle w, s' \rangle \). If \( V_{A,w} \) is empty, then obviously both these are trivially true. If \( V_{A,w} \) is non-empty, then the first of these is true just in case there is a world \( w' \) in the information state \( s \) relative to which we evaluate the complement of \( \forall A V's \) such that \( \forall p \) or \( \forall q \) is true at \( \langle s, w' \rangle \). But then either \( p \) or \( q \) is true at \( \langle s, w' \rangle \), in which case either \( \forall A V's \) might \( p \) or \( \forall A V's \) might \( q \) will be true at \( \langle w, s' \rangle \), in which case their disjunction will be as well.

\(^{38}\)A countermodel: let \( V_{A,w} = \{w, w'\} \), with \( f(w) = \{w\} \) and \( f(w') = \{w'\} \). Let \( p \) be true at \( w \) and false at \( w' \). Then \( \forall A V's \) might \( (p \) or \( \forall \; \neg p ) \) is true at \( \langle f, w \rangle \), since both worlds in \( V_{A,w} \) can access \( p \) \( \forall \; \neg p \)-world under \( f \); but \( \forall (A V's \) might \( p \) or \( A V's \) might \( \forall \; \neg p ) \) is false, since it is not the case that both worlds in \( V_{A,w} \) can access \( p \)-worlds, nor can both worlds in \( V_{A,w} \) access \( \forall \; \neg p \)-worlds. Whether the domain framework’s entailment is one we want to validate is not something I’ll try to settle; it is prima facie attractive, but it seems open that it is a pragmatic inference.
is strictly richer than the domain framework with respect to the embedding operators that can be defined in it.

Given the degree to which the subset relational system improves over the simple shifty system, it is tempting to conclude that the behavior of modals under attitudes in fact shows that we need the expressive power of the relational semantics. But this would be too fast. Just as we have seen that there are substantially weaker ways to account for Yalcin's data within the relational semantics than with Ninan's simple shifty system, likewise, there are substantially weaker ways to account for Yalcin's data within the domain semantics than with Yalcin's own simple shifty system. And, just as we explored the weakest possible way to account for Yalcin's data in the relational semantics, we should likewise explore the weakest possible constraint which accounts for those data in the domain semantics, asking whether such a constraint provides a plausible account of attitude predicates within the domain framework.

In doing so, I will assume, again, Boolean semantics for the connectives, and I will assume that attitude states are to be represented as sets of possible worlds. Given these assumptions, the weakest constraint which ensures that \( \neg A \text{ supposes } (p \text{ and might not } p) \) entails that A's suppositions are inconsistent is the following: a modal in the complement of \( \neg A \text{ supposes} \) is always evaluated relative to a subset of A's supposition worlds. We can encode this constraint, which I call the subset domain constraint, as follows:

**Definition 4.6.2. Subset domain semantics:**

- \( [A \text{ V's } p]^{s,w} \)
  
  a. defined only if \( s \subseteq V_{A,w} \)  
  
  b. where defined, true iff \( \forall w' \in V_{A,w} : [p]^{s,w'} = 1 \)  

Note first that, given our background assumptions, the subset domain constraint suffices to account for Yalcin's data, i.e. to ensure that \( \neg A \text{ supposes } (p \text{ and might not } p) \) always entails that A's

---

39 Things are otherwise for the update or state-based semantics, which are locked into the simply shifty system, at least as long as they want to model attitudes as sets of worlds.

40 A different route to an approach which ends up being essentially equivalent to the subset domain semantics goes by adding an ordering source to Hacquard (2010)'s event-relative semantics; see Hacquard 2010, §6.1.2 for a brief discussion of this possibility.
suppositions are inconsistent. Suppose \( \Gamma A \) supposes \((p \text{ and might not } p) \)\(^\top\) is true in this framework.

Suppose next that A's suppositions are consistent. It follows from this semantics that (i) for every world \( w' \) in A’s supposition state, \( p \) is true at \( \langle s, w' \rangle \); and that (ii) for some world \( w'' \in s \), \( p \) is false at \( \langle s, w'' \rangle \). But we know by the subset domain constraint that \( s \) is a subset of A’s supposition state, and so it follows that, for some world \( w'' \) in A’s supposition state, \( p \) is false at \( \langle s, w'' \rangle \), contrary to (i). It follows that A's suppositions are not, after all, consistent. Thus \( \Gamma A \) supposes \((p \text{ and might not } p) \)\(^\top\) entails that A's suppositions are inconsistent. (This explanation extends straightforwardly to order and scope variants, as well as other attitude predicates.)

Note next that, given our background assumptions, the subset domain constraint is necessary to account for Yalcin’s data, i.e. to predict that \( \Gamma A \) supposes \((p \text{ and might not } p) \)\(^\top\) always entails that A's suppositions are inconsistent. Suppose that A’s suppositions are consistent, and that the subset domain constraint is violated, i.e. \( s \setminus S_{A, w} \neq \emptyset \). Let \( p \) denote \( S_{A, w} \). Then \( \Gamma A \) supposes \((p \text{ and might not } p) \)\(^\top\) will be true at \( \langle s, w \rangle \), since (i) all of A’s supposition world are \( p \)-worlds; and (ii) some world in \( s \) is a \( \bar{p} \)-world, since \( s \setminus S_{A, w} \) is non-empty and by construction includes only \( \bar{p} \)-worlds. So, whenever the subset domain constraint is violated, we will be able to find a value for \( p \) such that \( \Gamma A \) supposes \((p \text{ and might not } p) \)\(^\top\) is true even if A’s suppositions are consistent.

Thus the subset domain constraint is the weakest constraint which suffices to capture Yalcin’s data within a domain semantics (given our background assumptions). And, just as the subset relational system was an improvement over the simple shifty system, so the subset domain system is an improvement over the simple shifty system (though somewhat less substantially so).

To see this, let’s go through, again, the three problems I pointed to above for the simple shifty system. First, unlike the simple shifty approach, the present approach does not predict that non-modal \( p \) entails \( \Gamma A \) knows might \( p \)\(^\top\): factivity ensures that \( w \) is always included in \( K_{A, w} \), but not always included in the domain of quantification for modals under \( \Gamma A \) knows\(^\top\).\(^{41}\)

On the second problem, the subset domain system’s success is rather limited. The good news is that \( \Gamma A \) knows might \( p \)\(^\top\) entails \( \Gamma A \) believes might \( p \)\(^\top\): if the former is true at \( \langle s, w \rangle \), then \( s \) must

\(^{41}\)A countermodel: suppose \( p \) is true at \( w \) and false at \( w' \); suppose \( K_{A, w} = \{ w, w' \} \), and let \( s = \{ w' \} \). Then \( p \) is true at \( w \), but \( \Gamma A \) knows might \( p \)\(^\top\) is false at \( \langle s, w \rangle \), since it is not the case that, for every world in \( K_{A, w} \), \( p \) is true at some world in \( s \).
contain a p-world, and so "A believes might p" will be true at \( (s, w) \), no matter what A's beliefs at \( w \) are like. The bad news is that the converse direction is (almost) valid, just as for the simple shifty system. Suppose "A believes might p" is true at \( (s, w) \). Then, as long as A's beliefs are consistent at \( w \), "A knows might p" is predicted to be true at \( (s, w) \) as well, since, if A's beliefs are consistent at \( w \), the only way for "A believes might p" to be true at \( w \) is for \( s \) to contain a p-world, in which case "A knows might p" will be true at \( (s, w) \). In other words, the subset domain semantics, like the simply shifty semantics, predicts the inference from "A believes might p" and A's beliefs are consistent to "A knows might p" to be valid. This illustrates a place where the flexibility of the more expressive relational semantics might be exactly what we need to account for the data; although there is some room for maneuver on behalf of the domain semantics, the less expressive domain framework, when spelled out in the simplest possible way needed to account for Yalcin's data, validates an invalid inference.

Finally, the domain semantics predicts that "A V's must p" is not equivalent to "A V's p". Thus we do not have "A V's must p" equivalent to "A V's p". We do still have, however, that "A V's p" entails "A V's must p". It is not entirely clear to me whether this latter entailment is correct, but, regardless, it is clear that this is an improvement on the simple shifty system.

Because of its predictions about the relation between knowledge and belief, the subset domain system is somewhat less attractive than the subset relational system, but it is still a substantial improvement over the simple shifty approach—substantial enough that, it seems to me, we cannot rule out the domain system simply on the basis of its predictions about modals under attitudes. This is especially so given the simplicity considerations in its favor that result from the expressibility results proven above.

The choice between these systems will ultimately depend on further exploration of data involving embedded modals—including data like those discussed in the first chapter. There I argued that there were indeed data—namely, Wittgenstein disjunctions—which a domain or update semantics could not make sense of in a principled way, but which a broadly relational semantics could make

---

42 For instance, suppose that A's V-attitude worlds include both p- and q-worlds, and let \( s = V_{A,w} \cap p \). Then at \( (s, w) \), "A V's must p" will be true, while "A V's p" will be false.

43 If \( p \) is true at \( (w', s) \) for all \( w' \in V_{A,w} \), then \( p \) will be true at \( (w'', s) \) for all \( w'' \in s \), since \( s \subseteq V_{A,w} \).
sense of (the semantics I defended there is, like the standard relational semantics, strictly more expressive than the domain or update semantics).\textsuperscript{44}

I will not revisit those data now, however. My goal here has been, rather, to bring out the kinds of considerations that could tell in favor of one approach over another, given the expressibility results proven above; and in particular to illustrate the advantages of methodological parsimony in exploring the choice between two semantic frameworks. When faced with new data which a given framework fails to make sense of, we should always explore the weakest constraints which need to be added to that framework to make sense of those data, and then see if that constraint gives us a satisfying picture across the board. This approach will help us determine whether that framework is capable of making sense of those data in a broadly plausible fashion—or whether, by contrast, the data tell decisively against such a framework. In the present case, this parsimonious approach showed that—contrary to what one might think looking at Yalcin’s and Ninan’s simple shifty frameworks—there is a good chance the we can make sense of the behavior of modals under attitudes in the relational framework, and a (somewhat less good) chance of doing so in the domain framework as well.

4.7 Conclusion

I have argued that we can gain new insight into the debate about the meaning of epistemic modals by taking an abstract perspective on different theories of their meaning. I have shown that many of the revisionary theories which have been proposed in recent years are strictly less expressive than the standard relational theory, in the sense that, for any embedding operator which can be defined in those revisionary theories, a corresponding operator (an operator with exactly the same logic) can be defined in the relational theory. But the converse does not hold. This shows that the dialectic in this debate is roughly the opposite of what it is often taken to be. On the one hand, there should be some presumption in favor of those more restricted theories, as long as they can make sense of the relevant embedding data. On the other hand, there may be embedding data which show that we

\textsuperscript{44}The situation is subtle, however, since the state-based semantics—which is strictly less expressive than any of these semantics—can indeed account for Wittgenstein disjunctions. So the considerations raised by those data may not really be about expressibility but rather about more subtle issues involving explanatory adequacy.
must abandon one of the revisionary theories in favor of the relational theory, but not (within certain mild bounds) *vice versa*.  

In the last part of the chapter, I explored the choice between the relational and domain frameworks in particular, focusing on the behavior of modals under attitudes. I emphasized the importance of methodological parsimony in choosing between these frameworks: by finding the weakest way of making sense of Yalcin's data in each framework, we showed that we can do so without committing to the implausible simple shifty approach to attitudes. My discussion of the choice between the domain and relational frameworks, based just on the behavior of modals under attitudes, was somewhat inconclusive, though I suggested there is some reason to prefer the relational approach. Deciding between these frameworks will thus ultimately depend on further embedding behavior.  

I have focused on epistemic modals in this chapter because they provide an apt illustration of the utility of the kinds of expressive comparisons I have developed. While I hope this discussion has advanced our understanding of the meaning of epistemic modals, my broader goal is to develop a formal framework which I hope will have widespread application in semantics. Semantic theory has often advanced thanks to results regarding expressive power—for instance in the theory of tense and temporal adverbs, or of generalized quantifiers. There is much more work to do in exploring the applications of this framework, as well as exploring and extending the underlying formalism. Comparisons of relative potential expressive power cannot on their own determine whether a semantic theory is correct, but they can greatly clarify the dialectical relationship in which competing theories stand, and thus what kinds of evidence we can expect to find for and against them.

---

45 See (Kamp, 1970; Cresswell, 1990).
46 See Peters and Westerståhl (2008, Parts 3-4) for an overview.
47 Of particular interest is a set of questions about computational complexity: if $\mathcal{M} \preceq \mathcal{M}'$, that means that, for any operator we can give a semantics for in $\mathcal{M}$, we can replicate that operator in an extension of $\mathcal{M}'$, but this does not tell us anything about the relationship between the complexity of the semantics in $\mathcal{M}'$ to the semantics in $\mathcal{M}$. Ideally, we would like to know whether we can replicate all operators from possible extensions of $\mathcal{M}$ without an upgrade in computational complexity.
Appendices
Appendix A

The bounded theory of epistemic modality

With $g$ a variable assignment and $\kappa$ a context (set of worlds), the local relational semantics for epistemic modals runs:

(1) \([\text{Might}_i p]^{g,\kappa,w}\]
   
a. defined only if $\forall w': g(i)(w') \subseteq \kappa$;
   
b. if defined, true iff $\exists w' \in g(i)(w) : [p]^{g,\kappa,w'} = 1$

The semantics for connectives and quantifiers that follow from our local context algorithm, plus standard semantic assumptions about the underlying "classical" semantic value, run as follows:

(2) \([p \text{ and } q]^{g,\kappa,w} = 1\) iff $[p]^{g,\kappa_p,w} = 1$ and $[q]^{g,\kappa_q,w} = 1$

Where for any sentence $p$, variable assignment $g$, and context $\kappa$: $\kappa_p = \kappa \cap \{w : [p]^{g,\kappa,w} = 1\}$.

(3) \([p \text{ or } q]^{g,\kappa,w} = 1\) iff $[p]^{g,\kappa_p,w} = 1$ or $[q]^{g,\kappa_q,w} = 1$

Where for any sentence $p$, variable assignment $g$, and context $\kappa$: $\kappa_p = \kappa \setminus \{w : [p]^{g,\kappa,w} = 1\}$.

(4) \([\text{Not } p]^{g,\kappa,w} = 1\) iff $[p]^{g,\kappa,w} = 0$

(5) \([A \text{ supposes } q]^{g,\kappa,w} = 1\) iff $\forall w' \in S_{A,w} : [q]^{g,\bigcup_{w' \in \kappa} S_{A,w,w'}} = 1$

Where for any agent $A$ and world $w$, $S_{A,w}$ the worlds compatible with $A$'s suppositions in $w$.

(6) \([\text{If } p, q]^{g,\kappa,w} = 1\) iff $[q]^{g,\kappa_q,w} = 1$\(^{1}\)

\(^{1}\)This implicitly assumes that $q$ always contains a modal, perhaps a tacit modal, which is sensitive to its local context. We may assume this modal is just epistemic 'must', as in Kratzer 1986, though we need not do so. The key observation for
(7) \([\exists x : [p]^g_{\kappa, \omega}(x) = 1, w(x) = 1, [q]^g_{\kappa, \omega}(x) = 1, w(x) = 1]\)

(8) \([\exists x : [p]^g_{\kappa, \omega}(x) = 1] \iff [q]^g_{\kappa, \omega}(x) = 1\)

(9) \([\forall x : [p]^g_{\kappa, \omega}(x) = 1] \implies [q]^g_{\kappa, \omega}(x) = 1\)

present purposes is, again, that \(\neg p \land \neg p, \then \ldots \) will be equivalent to \(\neg p \land p, \then \ldots \)
Appendix B

The contextualist framework

Here I give an explicit model of the contextualist theory of communication, and verify that this model, together with the prospective contextualist account of the assertoric content of modal claims developed in the second chapter, predicts that an assertion of "Might \( p \)" is a proposal to make \( p \) compatible with the common ground, and an assertion of "Must \( p \)" a proposal to make \( p \) entailed by the common ground. What follows is just one possible implementation of the contextualist model; subtle questions of implementation, however, do not matter for present purposes.$^1$

We can model contextualism’s common ground at \( t \) with a consistent, logically closed set of sentences \( \Gamma_t \) in a standard modal language. As above, I use \( \square_t \) as a modal operator interpreted as ‘it is common ground at \( t \) that’, and \( \lozenge_t \) as its dual. We read off the common ground at a time \( t \) as follows: for any sentence \( p \), \( p \) is common ground at \( t \) iff \( \square_t p \in \Gamma_t \). In order to encode the monotonicity assumption implicit in the contextualist model—the assumption that conversants proceed as if they will gain, rather than lose, information as conversation goes on—we stipulate that whenever \( \square_t' p \) is in \( \Gamma_t' \), so is \( \square_t p \) for all \( t'' \) after \( t' \).

With this background, we can say that, in the contextualist framework, an assertion of \( p \) at time \( t \) is a proposal which, if accepted, has the effect of ensuring that \( \square_{t'} p \) is in \( \Gamma_{t'} \), where \( t' \) is the time at which \( p \) is either accepted or rejected (i.e. the prospective time, relative to that assertion). We further stipulate that \( \Gamma_{t'} \) is the smallest set which includes \( \Gamma_t \), updated with the fact that the assertion in question has been made and anything that follows from this (i.e. with any information which is accommodated; see Stalnaker 2002; von Fintel 2008); which includes \( \square_{t'} p \); \( \square_{t''} p \) for all \( t'' \) after \( t' \); and is logically closed. Some revision of the common ground may be necessary to ensure that this set is consistent (as when, for instance, \( p \) is common ground and then "Not \( p \)" or "Might

---

$^1$The contextualist framework is sometimes glossed as an intersective update model: when \( p \) is asserted at a context set \( c \), the subsequent context set is \( c \cap p \). This is a reasonable approximation, but it ignores a feature of updating that is crucial to our present purposes, and that is built into the orthodox framework: namely, that when one updates with \( p \), one also updates with the proposition that \( p \) is commonly accepted, and so on; these updates are critical so that the logic of common ground is preserved across updates. This is accounted for in the present, slightly more complicated, implementation of contextualism.
not \textit{p}^\neg is asserted); I assume this revision comes in in the accommodation step. The mechanics of how such a revision goes is beyond our scope here: although this topic is of special interest for the theory of epistemic modals, since assertions of epistemic possibility claims are a standard tactic for creating such a clash, the underlying theory of how such a revision goes is just an instance of the general problem of belief revision, and need not be addressed by our theory of modals.\textsuperscript{2} We assume, finally, that if an assertion is rejected (in the sense we are interested here), this is equivalent to its negation being accepted.

Because of the meaning assigned to modal claims by prospective contextualism, we can immediately regiment them in our formal language: ‘might’ and ‘must’ will just be treated as \( \diamond_{t'} \) and \( \Box_{t'} \), respectively, where \( t' \) is the prospective time. Thus our logic for modal claims—and our formal model for updates with them—will follow immediately from our model for updates in general. A claim of ‘Might \textit{p}’ as asserted at \( t \) will thus be regimented as \( \diamond_{t'} \textit{p} \), where \( t' \), again, is the prospective time relative to the assertion (again, what time that actually amounts to will depend on concrete features of the actual speech situation). Suppose first that this claim is accepted. Then \( \Box_{t'} \diamond_{t'} \textit{p} \) is in \( \Gamma_{t'} \). As proven in §2.3, it is a theorem of the logic of common ground that \( \Box_{t'} \diamond_{t'} \textit{p} \rightarrow \diamond_{t'} \textit{p} \). From this and the fact that \( \Gamma_{t'} \) is logically closed, it follows that \( \diamond_{t'} \textit{p} \in \Gamma_{t'} \). Under the assumption that \( \Gamma_{t'} \) is always consistent (assuming that some tacit adjustment happens when this would otherwise not be the case), it follows that the \( \Box_{t'} \neg \textit{p} \notin \Gamma_{t'} \), and thus \( \neg \textit{p} \) is not common ground at \( t' \), and thus that \( \textit{p} \) is compatible with the common ground at \( t' \).

Suppose second that this claim is rejected. Then \( \Box_{t'} \neg \diamond_{t'} \textit{p} \) is in \( \Gamma_{t'} \); equivalently, \( \Box_{t'} \Box_{t'} \neg \textit{p} \in \Gamma_{t'} \). As proven in §2.3, it is a theorem in the logic of common ground entails that \( \Box_{t'} \Box_{t'} \textit{p} \rightarrow \Box_{t'} \textit{p} \). From this and the fact that \( \Gamma_{t'} \) is logically closed, it follows that \( \Box_{t'} \neg \textit{p} \in \Gamma_{t'} \), and thus that \( \neg \textit{p} \) is common ground at \( t' \), and \( \textit{p} \) is not compatible with the common ground at \( t' \).

Since ‘Must \textit{p}’ will be regimented as \( \Box_{t'} \textit{p} \), these considerations also show that an assertion of ‘Must \textit{p}’ is a proposal which, if accepted, ensures that \( \textit{p} \) is in the common ground; and, if rejected, ensures that \( \neg \textit{p} \) is compatible with the common ground.

\textsuperscript{2}See C.E Alchourrón and P. Gärdenfors and D. Makinson 1985 for a classic reference.
Appendix C

Definitions and proofs

In this appendix I give definitions of the technical terms used in §4.3, and provide proofs of the claims made there.

C.1 Definitions

**Definition C.1.1.** *Model:* Given a propositional language \( \mathcal{L} \), built from a vocabulary comprising a set \( \mathcal{A} \) of atomic sentences and sentence operators \( \mathcal{O} \), and comprising all and only (i) atoms from \( \mathcal{A} \), and (ii) strings of the form \( O^n(\varphi_1, \varphi_2, \ldots, \varphi_n) \) for any \( n \)-place sentence operator \( O^n \in \mathcal{O} \), and sentences \( \varphi_i : 1 \leq i \leq n \) in \( \mathcal{L} \), a model of \( \mathcal{L} \) is a sequence \( \langle \mathcal{W}, \mathcal{I}, \nu, [] \rangle \). \( \mathcal{W} \) is a set of possible worlds, and \( \mathcal{I} \) is a set of indices; formally speaking, these can be any non-empty sets. \( \nu \) is an atomic valuation function, which takes any atomic sentence of \( \mathcal{L} \) and any possible world from \( \mathcal{W} \) to either 1 or 0. \( [] \) is an interpretation function, which takes an atomic sentence to a set of indices; takes an \( n \)-place sentence operator \( O^n \) to a function from an \( n \)-tuple of sets of indices to a set of indices; and take any sentence of the form \( O^n(\varphi_1, \varphi_2, \ldots, \varphi_n) \), for any \( n \)-place sentence operator \( O^n \in \mathcal{O} \) and any \( n \)-tuple \( \langle \varphi_1, \varphi_2, \ldots, \varphi_n \rangle \) of sentences of \( \mathcal{L} \), to \( [O^n]\langle [\varphi_1], [\varphi_2], \ldots, [\varphi_n] \rangle \); and takes any sentence \( \varphi \) of \( \mathcal{L} \) and any index \( i \) (written \([\varphi]^i\)) to 1 ("true") just in case \( i \in [\varphi] \), and otherwise to 0 ("false").

Given a quantified language \( \mathcal{L}^* \), built from a vocabulary comprising a set \( \mathcal{V} \) of variables, a set \( \mathcal{R} \) of relation symbols, and a set \( \mathcal{O} \) of sentence operators; and comprising all and only (i) atoms of the form \( R^n(\langle x_1, x_2, \ldots, x_n \rangle) \), for \( R^n \) an \( n \)-place relation in \( \mathcal{R} \), and \( x_i : 1 \leq i \leq n \) variables from \( \mathcal{V} \); and (ii) strings of the form \( O^n(\langle \varphi_1, \varphi_2, \ldots, \varphi_n \rangle) \) for any \( n \)-place sentence operator \( O^n \in \mathcal{O} \) and any sentences \( \varphi_i : 1 \leq i \leq n \) in \( \mathcal{L}^* \), a model of \( \mathcal{L}^* \) is a sequence \( \langle D, \mathcal{W}, \mathcal{I}, \nu, [] \rangle \). \( D \) is a domain of individuals, \( \mathcal{W} \) is a set of possible worlds, and \( \mathcal{I} \) is a set of indices. From a formal point of view,

---

1 We will generally use lower-case italic letters to range over atoms, and Greek letters to range over all sentences.
2 Model of" is often used to describe a model which makes true every sentence in a given set of sentences; that is not how I am using it here.
these can, again, be any non-empty sets. \( \nu \) is an atomic valuation function, which takes a possible world and an \( n \)-place relation symbol to an \( n \)-place relation (a subset of \( D^n \), the set of \( n \)-tuples of elements of \( D \)). \([ \cdot ]\) is an interpretation function, defined exactly as above.

**Definition C.1.2. Extension of a Language:** Given a language \( \mathcal{L} \), and a set \( \mathcal{O} \) of sentence operators disjoint from the vocabulary of \( \mathcal{L} \), the extension of \( \mathcal{L} \) to \( \mathcal{O} \), written \( \mathcal{L}^\mathcal{O} \), is the smallest set containing \( \mathcal{L} \) and closed under the elements of \( \mathcal{O} \), i.e. where \( a \) is the function giving the arity of sentence operators, the smallest set containing \( \mathcal{L} \) and \( \{ \psi : \exists O \in \mathcal{O} : \exists n : a(O) = n \land \exists I \subseteq \mathcal{L}^n : \psi = O^n(I) \} \).

**Definition C.1.3. Extension of a Model:** Given a language \( \mathcal{L} \), a model \( \mathcal{M} \) of \( \mathcal{L} \), and an extension \( \mathcal{L}^\mathcal{O} \) of \( \mathcal{L} \), an extension of \( \mathcal{M} \) to a model of \( \mathcal{L}^\mathcal{O} \) with respect to \( \mathcal{L} \) is any model \( \mathcal{M}' \) of \( \mathcal{L}^\mathcal{O} \) which has the same domain of individuals as \( \mathcal{M} \) (if \( \mathcal{M} \) has a domain), the same set of possible worlds, the same set of indices, and the same atomic valuation as \( \mathcal{M} \); and which is such that for any \( \varphi : \varphi \in \mathcal{L} \rightarrow [\varphi]_{\mathcal{M}^\mathcal{O}} = [\varphi]_{\mathcal{M}} \).

### C.2 Proofs

For convenience, I repeat the definition of relative potential expressibility here; I then turn to proofs of the claims of §4.3:

**Definition 4.3.1. Relative Potential Expressibility:** For any models \( \mathcal{M} \) and \( \mathcal{M}' \) of a language \( \mathcal{L} \), \( \mathcal{M} \preceq \mathcal{M}' \) iff, for any set of new operators \( \mathcal{O} \), for any extension \( \mathcal{M}^\mathcal{O} \) of \( \mathcal{M} \) to \( \mathcal{L}^\mathcal{O} \), there is an extension \( \mathcal{M}'^\mathcal{O} \) of \( \mathcal{M}' \) to \( \mathcal{L}^\mathcal{O} \) which preserves the logic of \( \mathcal{O} \) from \( \mathcal{M}^\mathcal{O} \): that is, which is such that, for any \( O^n \in \mathcal{O} \), for any sentence \( \psi \) in \( \mathcal{L}^\mathcal{O} \) and \( n \)-tuple \( \varphi \) of sentences from \( \mathcal{L}^\mathcal{O} \), \((O^n(\varphi) \vdash_{\mathcal{M}^\mathcal{O}} \psi) \leftrightarrow (O^n(\varphi) \vdash_{\mathcal{M}'^\mathcal{O}} \psi) \). For a set of sentences \( \Phi \), sentence \( \psi \), and model \( \mathcal{M} \), \( \Phi \vdash_{\mathcal{M}} \psi \) iff \( \psi \) is true in \( \mathcal{M} \) at every index where all the elements of \( \Phi \) are true in \( \mathcal{M} \).

**Fact 4.3.1. Characterization of Expressibility:** For any models \( \mathcal{M} \) and \( \mathcal{M}' \) of a language \( \mathcal{L} \), \( \mathcal{M} \preceq \mathcal{M}' \) just in case there is a function \( g \) (call it a *witness function*) from the indices of \( \mathcal{M} \) to those of \( \mathcal{M}' \) which is such that (i) for any sentence \( \varphi \) of \( \mathcal{L} \) and index \( i \) in \( \mathcal{M} \), \( \varphi \) is true at \( i \) in \( \mathcal{M} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}' \); and (ii) \( g \) is an injection.

The proof of Fact 4.3.1 goes by way of a lemma:

**Lemma C.2.1.** For any models \( \mathcal{M} \) and \( \mathcal{M}' \) of a language \( \mathcal{L} \), \( \mathcal{M} \preceq \mathcal{M}' \) iff for any extension \( \mathcal{L}^\mathcal{O} \) of \( \mathcal{L} \) with a set of new operators, and any extension \( \mathcal{M}^\mathcal{O} \) of \( \mathcal{M} \) to a model of \( \mathcal{L}^\mathcal{O} \), there is an extension

---

3 I will call any set of operators which meets this novelty constraint a "set of new operators"; I sometimes leave this novelty condition implicit in introducing extensions of languages.

4 I will usually leave the relativization to the initial language implicit. For any model \( \mathcal{M} \), I write \([ \cdot ]_{\mathcal{M}} \) for \( \mathcal{M} \)'s interpretation function.

186
$M'^O$ of $M'$ to $L'^O$ such that there exists a function $g$ from the indices of $M$ to the indices of $M'$ such that for any sentence $\varphi$ of $L'^O$ and index $i$ in $M$, $\varphi$ is true at $i$ in $M'^O$ iff $\varphi$ is true at $g(i)$ in $M'^O$.

Proof. [$\Rightarrow$] For arbitrary models $M$ and $M'$ of an arbitrary language $L$, suppose $M \preceq_L M'$. Recall that means that, for any set of new operators $O$ and any extension $M'^O$ of $M$ to $L'^O$, there is an extension $M'^O$ of $M'$ to $L'^O$ which preserves the logic of $O$ from $M'^O$. Consider an arbitrary set of new operators $O$ and arbitrary extension $M'^O$ of $M$ to $L'^O$. We will show that there is a function $g$ from the indices of $M$ to the indices of $M'$ such that for any sentence $\varphi$ of $L'^O$ and index $i$ in $M$, $\varphi$ is true at $i$ in $M'^O$ iff $\varphi$ is true at $g(i)$ in $M'^O$.

We do so by way of considering first a different extended language which contains $L'^O$, and a different extension of $M$ which also extends $M'^O$. In particular, consider the extension of $L$ to $O \cup N$, where $O \cap N = \emptyset$, no operator in $N$ is in the vocabulary of $L$, and the cardinality of $N$ is greater than or equal to the cardinality of the set of indices of $M$. Now extend $M$ to a new model $M'^O_N$ of the resulting language, $L'^O \cup N$, with the following properties:

(a) $M'^O_N$ is an extension of $M'^O$, so that $\forall \varphi \in L'^O: [[\varphi]]_{M'^O} = [[\varphi]]_{M'^O_N};$

(b) for some sentence $\psi$, for each index $i$ of $M$, there is an $O \in N$, call it $O_i$, which uniquely specifies $i$, in the sense that $O_i(\psi)$ is true at $i$ in $M'^O_N$ and false everywhere else in $M'^O_N$; and

(c) for some unary sentence operator $\neg \in N$, $\neg$ is given the classical semantics of negation in $M'^O_N$, i.e. for any $\varphi \in L'^O \cup N$, $\neg(\varphi)$ is true at $i$ iff $\varphi$ is not true at $i$.

It is easy to see that this will always be possible.

Now extend $M'$ to a model $M'^O_N$ of $L'^O \cup N$, which preserves the logic of $O \cup N$ from $M'^O_N$; we know this will be possible by our assumption that $M \preceq_L M'$. Now, define a function $g$ such that, for any index $i$ in $M'^O_N$, $g$ takes $i$ to an index $g(i)$ in $M'^O_N$ such that (i) $O_i(\psi)$ is true at $g(i)$ in $M'^O_N$, and (ii) some sentence in $L'^O \cup N$ is false at $g(i)$. We know there is such an index; otherwise, we would have $O_i(\psi) \models_{M'^O_N} \neg(\psi)$, and thus $O_i(\psi) \models_{M'^O_N} \neg(O_i(\psi))$, but we know the latter is false by our semantics for $O_i$ and $\neg$ in $M'^O_N$.

Now for any sentence $\varphi$ of $L'^O \cup N$:

- Suppose first $\varphi$ is true at $i$ in $M'^O_N$. Then $O_i(\psi) \models_{M'^O_N} \varphi$, and thus by our assumption that $O_i$ has the same logic in $M'^O_N$ as in $M'^O$, $O_i(\psi) \models_{M'^O} \varphi$, and thus $\varphi$ is true at $g(i)$, since $O_i(\psi)$ is true at $g(i)$.

- Suppose next that $\varphi$ is not true at $i$ in $M'^O_N$. Then $\neg \varphi$ is true at $i$ in $M'^O_N$, and thus $O_i(\psi) \models_{M'^O_N} \neg \varphi$ and thus $O_i(\psi) \models_{M'^O_N} \neg \varphi$, and thus $\neg \varphi$ is true at $g(i)$ in $M'^O_N$, since $O_i(\psi)$ is true at $g(i)$. Thus we can conclude that $\varphi$ is not true at $g(i)$ in $M'^O_N$; if it were, since our logic for negation is classical in $M'^O_N$, and thus in $M'^O_N$, we would have that everything is true at $g(i)$ in $M'^O_N$, contrary to assumption.
Thus for any \( \varphi \in \mathcal{L}^{O \cup N} \), we have \( \varphi \) true at \( i \) in \( \mathcal{M}_N^{O} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}_N^{O} \); thus in particular, for any \( \varphi \in \mathcal{L}^{O} \), which is a subset of \( \mathcal{L}^{O \cup N} \), \( \varphi \) is true at \( i \) in \( \mathcal{M}_N^{O} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}_N^{O} \); and, since \( \mathcal{M}_N^{O} \) is an extension of \( \mathcal{M}^{O} \), it follows that for any \( \varphi \in \mathcal{L}^{O} \), \( \varphi \) is true at \( i \) in \( \mathcal{M}^{O} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}^{O} \). Since \( O \) and \( \mathcal{M}^{O} \) were selected arbitrarily, this shows that, for any extension \( \mathcal{L}^{O} \) of \( \mathcal{L} \) and extension \( \mathcal{M}^{O} \) of \( \mathcal{M} \), we can find an extension of \( \mathcal{M}' \) to \( \mathcal{L}^{O} \) with the property that there is a function from the indices of \( \mathcal{M} \) to those of \( \mathcal{M}' \) which preserves truth for the sentences of \( \mathcal{L}^{O} \) in the extended models.

### [\( \Leftrightarrow \)]

For arbitrary models \( \mathcal{M} \) and \( \mathcal{M}' \) of an arbitrary language \( \mathcal{L} \), suppose that, for an arbitrary extension \( \mathcal{L}^{O} \) of \( \mathcal{L} \) with a set of sentence operators, and an arbitrary extension \( \mathcal{M}^{O} \) of \( \mathcal{M} \) to a model of \( \mathcal{L}^{O} \), there is always an extension \( \mathcal{M}'^{O} \) of \( \mathcal{M}' \) to \( \mathcal{L}^{O} \) such that there exists a function \( g \) from the indices of \( \mathcal{M} \) to the indices of \( \mathcal{M}' \) such that for any sentence \( \varphi \in \mathcal{L}^{O} \) and index \( i \) in \( \mathcal{M} \), \( \varphi \) is true at \( i \) in \( \mathcal{M}^{O} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}'^{O} \). We can use this fact to construct a new extension \( \mathcal{M}'^{O-} \) of \( \mathcal{M}' \) which matches the logic of \( O \) in \( \mathcal{M} \), as follows. Let \( \mathcal{M}'^{O-} \) be just like \( \mathcal{M}'^{O} \), except that, at every index \( i \) of \( \mathcal{M}'^{O} \) which is not in the image of \( g \), for any sentence \( \varphi \in \mathcal{L}^{O} \setminus \mathcal{L} \), \( \varphi \) is false at \( i \) in \( \mathcal{M}'^{O-} \). Note that \( \mathcal{M}'^{O-} \) is still an extension of \( \mathcal{M}' \) to \( \mathcal{L}^{O} \); and \( g \) will still preserve truth for the relevant sentences: since we did not change the truth of any sentences in the image of \( g \), for any \( \varphi \in \mathcal{L}^{O} \setminus \mathcal{L} \), \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}'^{O-} \).

We now turn to the proof of Fact 4.3.1:

### Proof:

\( \Rightarrow \) Suppose for arbitrary \( \mathcal{M}, \mathcal{M}' \) and \( \mathcal{L} \), there is no function \( g \) from the indices of \( \mathcal{M} \) to those of \( \mathcal{M}' \) which is such that (i) for any sentence \( \varphi \) of \( \mathcal{L} \) and index \( i \) in \( \mathcal{M} \), \( \varphi \) is true at \( i \) in \( \mathcal{M} \) iff \( \varphi \) is true at \( g(i) \) in \( \mathcal{M}' \); and (ii) \( g \) is an injection. Find a set of new operators \( O \) with cardinality equal to the set of indices in \( \mathcal{M} \). Let \( f \) be a bijection from the indices of \( \mathcal{M} \) to \( O \). Extend \( \mathcal{M} \) to a new model \( \mathcal{M}^{O} \) of \( \mathcal{L}^{O} \), with the property that, for any index \( i \) in \( \mathcal{M} \), for any sentence \( \varphi \in \mathcal{L}^{O} \), \( f(i) \) is true \( \mathcal{M}^{O} \) at \( i \) and false \( \mathcal{M}^{O} \) at every other index of \( \mathcal{M} \); that is, \( f(i) \) "tags" \( i \) in \( \mathcal{M}^{O} \). Now

\[ \text{5 'True}_{\mathcal{M}'} \text{ is shorthand for 'true in } \mathcal{M}' \]
consider an arbitrary extension $\mathcal{M}'^O$ of $\mathcal{M}'$. Suppose there is a function $g$ from the indices of $\mathcal{M}$ to the indices of $\mathcal{M}'$ with the property that, for all $\varphi \in \mathcal{L}'$, $\varphi$ is true in $\mathcal{M}$ at $i$ iff $\varphi$ is true in $\mathcal{M}'^O$ at $g(i)$. Since $\mathcal{L} \subset \mathcal{L}'$, and since extensions of models of a given language preserve truth for sentences in the original language, we know that, for all $\varphi \in \mathcal{L}$, $\varphi$ is true in $\mathcal{M}$ at $i$ iff $\varphi$ is true in $\mathcal{M}'$ at $g(i)$. Then it follows from our assumption that $g$ is not an injection: for some $\mathcal{M}$-indices $i$ and $i'$ with $i \neq i'$, $g(i) = g(i')$. Choose some $\varphi \in \mathcal{L}'$. We know by construction of $\mathcal{M}'^O$ that $f(i)(\varphi)$ is true in $\mathcal{M}$ at $i$ and false in $\mathcal{M}'$ at $i'$. But, since $g(i) = g(i')$, $f(i)(\varphi)$ will either be true in $\mathcal{M}'^O$ at both $g(i)$ and $g(i')$, or false at both, and thus it will not be the case that, for every sentence $\varphi$ of $\mathcal{L}'$, if $\varphi$ is true at $i$ in $\mathcal{M}'$, then $\varphi$ is true at $g(i)$ in $\mathcal{M}'^O$, contrary to assumption. Thus, since $\mathcal{M}'^O$ was chosen arbitrarily, there is no function which preserves truth between $\mathcal{M}'^O$ and $\mathcal{M}'$; and thus by Lemma C.2.1, $\mathcal{M} \not\models \mathcal{L}'$. 

$[\models]$ Suppose, for arbitrary $\mathcal{M}$, $\mathcal{M}'$ and $\mathcal{L}$, there is such a truth-preserving injection $g$. Given an arbitrary set $\mathcal{O}$ of sentence operators and an arbitrary extension $\mathcal{M}'^O$ of $\mathcal{M}$ to $\mathcal{L}'$, we show there is an extension $\mathcal{M}'^O$ of $\mathcal{M}'$ to $\mathcal{L}'$ which has the property that, for any sentence $\varphi \in \mathcal{L}'$, $\varphi$ is true at an index $i$ in $\mathcal{M}'^O$ just in case $\varphi$ is true at $g(i)$ in $\mathcal{M}'^O$. Let $K$ index the elements of $\mathcal{O}$. For each $O_k: k \in K$, extend $\mathcal{M}$ to the model $\mathcal{M}_k$ which is just like $\mathcal{M}$, except its interpretation function $[\cdot]_{\mathcal{M}_k}$ is augmented with the semantic rule for $O_k$ from $[\cdot]_{\mathcal{M}'^O}$. Then, for each $O_k$, extend $\mathcal{M}'$ to the model $\mathcal{M}'_k$ which augments the interpretation function of $\mathcal{M}'$ with a semantic rule for $O_k$ as follows. For brevity, for any set $\alpha$ and function $f$, define $f[\alpha]$ to be the pointwise application of $f$ to $\alpha$ where defined, i.e. $f[\alpha] = \{f(a) : a \in \alpha \land f(a) \text{ is defined}\}$. Let $g^{-1}$ be the inverse of $g$, defined only on the image of $g$; that $g^{-1}$ is a well-defined function follows because $g$ is an injection (which in turn follows by assumption). Now, suppose first that $O_k$ is a unary sentence operator; then let $\mathcal{M}'_k$ extend $\mathcal{M}'$ with the following semantic rule: $[O_k]_{\mathcal{M}'_k} = \lambda s_{\mathcal{M}'}: g([O_k]_{\mathcal{M}_k}(g^{-1}[s]))$, where $s_{\mathcal{M}'}$ ranges over sets of $\mathcal{M}'$ indices. Thus in $\mathcal{M}'_k$, $O_k$ takes a set of $\mathcal{M}'$ indices; then finds the pre-image (where defined) of this complement with respect to $g$; then applies the semantic rule for $O_k$ in $\mathcal{M}_k$ to this pre-image; and finally, returns the pointwise application of $g$ to the resulting set. Now note that, for any set $s$ of $\mathcal{M}$-indices and set $s'$ of $\mathcal{M}'$-indices, if $i \in s \leftrightarrow g(i) \in s'$, it follows that $i \in [O_k]_{\mathcal{M}_k}(s) \leftrightarrow g(i) \in [O_k]_{\mathcal{M}'_k}(s')$. To see this, assume for arbitrary $s$ and $s'$ that $i \in s \leftrightarrow g(i) \in s'$. Now note that $s = g^{-1}[s']$: if $i \in s$, then by assumption $g(i) \in s'$, and thus $g^{-1}[s']$ will include $i$, by construction; and if $i \notin s$, then by assumption $g(i) \notin s'$, and since $g^{-1}$ is an injection, by construction, we know that $i \notin g^{-1}[s']$. We thus have $[O_k]_{\mathcal{M}'_k}(s') = g([O_k]_{\mathcal{M}_k}(g^{-1}[s'])) = g([O_k]_{\mathcal{M}_k}(s))$. In other words, whenever $i \in s \leftrightarrow g(i) \in s'$, then $[O_k]_{\mathcal{M}'_k}(s')$ is just the pointwise application of $g$ to $[O_k]_{\mathcal{M}_k}(s)$, and thus, since $g$ is an injection, $i \in [O_k]_{\mathcal{M}_k}(s) \leftrightarrow g(i) \in [O_k]_{\mathcal{M}'_k}(s')$. The generalization of this construction to $n$-place sentence operators, for any $n$, is straightforward. We use this method to construct $\mathcal{M}'_k$ for all $k \in K$. 

189
Now, where $M' = \langle W, I, v, [\cdot]_{M'} \rangle$ or $\langle D, W, I, v, [\cdot]_{M'} \rangle$, let $M'^{O} = \langle W, I, v, \bigcup_{k \in K} [\cdot]_{M'_k} \rangle$
or $\langle D, W, I, v, \bigcup_{k \in K} [\cdot]_{M'_k} \rangle$, respectively. By our construction, we know that for any $O \in \mathcal{O}$ and any sets of $M$-indices $s$ and $M'$-indices $s'$ such that $i \in s \leftrightarrow g(i) \in s'$, $[O]_{M'^{O}}(s') = g([O]_{M^{O}}(s))$, and thus $i \in [O]_{M'^{O}}(s') \leftrightarrow g(i) \in [O]_{M^{O}}(s')$. We know by assumption that, for all sentences $\varphi \in \mathcal{L}$, $i \in [\varphi]_{M} \leftrightarrow g(i) \in [\varphi]_{M'}$, and thus (since extending a model never changes its interpretation of a sentence already in the language of the original model) $i \in [\varphi]_{M'^{O}} \leftrightarrow g(i) \in [\varphi]_{M^{O}}$. Now consider any sequence of sentences $\tilde{\psi}$ with the property that for each $\psi_j$ in the sequence, $i \in [\psi_j]_{M'^{O}} \leftrightarrow g(i) \in [\psi_j]_{M^{O}}$. Then we know that, by our construction, for any $k \in K$ and index $i$ of $M$, $i \in [O_k(\tilde{\psi})]_{M'^{O}} \leftrightarrow g(i) \in [O_k(\tilde{\psi})]_{M^{O}}$. Since the sentences of $\mathcal{L}^{O}$ are built recursively from the sentences of $\mathcal{L}$ and the operators in $\mathcal{O}$, it follows by an induction on the complexity of formulae that, for any $\varphi \in \mathcal{L}^{O}$, $i \in [\varphi]_{M'^{O}} \leftrightarrow g(i) \in [\varphi]_{M^{O}}$. Since $\mathcal{O}$ and $M'^{O}$ were chosen arbitrarily, we conclude that, for any set of new operators $\mathcal{O}$, for any extension $M'^{O}$ of $M$ to $\mathcal{L}^{O}$, there is an extension $M'^{O}$ of $M'$ to $\mathcal{L}^{O}$ such that there is a function $g$ with the property that, for any sentence $\varphi \in \mathcal{L}^{O}$, $\varphi$ is true at an index $i$ in $M'^{O}$ just in case $\varphi$ is true at $g(i)$ in $M'^{O}$; and thus by Lemma C.2.1, $M \preceq_{\mathcal{L}} M'$.

\[\square\]

**Fact 4.3.2.** $\mathcal{D} \preceq_{\mathcal{L}^{O}} \mathcal{R}$.

**Proof:** We assume the same stock of worlds and atomic valuations for both models. Construct a function $g$ as follows. For any index $\langle s, w \rangle$ in $\mathcal{D}$, let $g(\langle s, w \rangle) = \langle f^s, w \rangle$, where $f^s$ is the constant function from worlds to $s$. For any atomic sentence $p$ of $\mathcal{L}^{O}$, $p$ will be true$_{\mathcal{D}}$ at $i$ iff $p$ is true$_{\mathcal{R}}$ at $g(i)$, since we are assuming the same stock of worlds and atomic valuation in both models, and since the truth of atomic sentence in these frameworks depends only on the world parameter of the index, and atomic valuation. Now for any sentence $\varphi \in \mathcal{L}^{O}$, assume for induction that $\varphi$ is true$_{\mathcal{D}}$ at $i$ iff it is true$_{\mathcal{R}}$ at $g(i)$. We show that, for arbitrary index $i$, $\Diamond \varphi$ is true$_{\mathcal{D}}$ at $i$ iff $\Diamond \varphi$ is true$_{\mathcal{R}}$ at $g(i)$. $i$ will have the form $\langle s, w \rangle$, for information state $s$ and world $w$, and, by our semantics for $\Diamond$ in $\mathcal{D}$, $\Diamond \varphi$ will be true$_{\mathcal{D}}$ at $i$ iff $\varphi$ is true$_{\mathcal{D}}$ at some element in the set $\Phi = \{ \langle s, w' \rangle : w' \in s \}$. $g(i)$ will have the form $\langle f^s, w \rangle$, and, by our semantics for $\Diamond$ in $\mathcal{R}$, $\Diamond \varphi$ will be true$_{\mathcal{R}}$ at $g(i)$ iff $\varphi$ is true$_{\mathcal{R}}$ for some element in the set $\Psi = \{ \langle f^s, w' \rangle : w' \in f^s(w) \}$. Now note that, thanks to the way we constructed $g$ and the fact that $f^s(w) = s$, $g$ will be a bijection from $\Phi$ to $\Psi$. And so it follows from our assumption for induction that $\varphi$ will be true$_{\mathcal{D}}$ at some element in $\Phi$ just in case $\varphi$ is true$_{\mathcal{R}}$ at some element in $\Psi$, and thus $\Diamond \varphi$ will be true$_{\mathcal{D}}$ at $i$ iff $\Diamond \varphi$ is true$_{\mathcal{R}}$ at $g(i)$. It thus follows by induction on the complexity of formulae that, for any sentence $\varphi$ of $\mathcal{L}^{O}$ and any index $i$, $\varphi$ is true$_{\mathcal{D}}$ at $i$ iff $\varphi$ is true$_{\mathcal{R}}$ at $g(i)$. Finally, it is easy to see that $g$ is an injection. Given Fact 4.3.1, it thus follows that $\mathcal{D} \preceq_{\mathcal{L}^{O}} \mathcal{R}$.

\[\square\]
Fact 4.3.4. $\mathcal{R} \not\preceq_{L^0} \mathcal{D}$.

Proof. Consider an $\mathcal{R}$-index $(f, w)$, with $f(w) = \{w'\}$, $v(p, w') = 0$, $f(w') = \{w''\}$, and $v(p, w'') = 1$. Then $\Diamond p$ will be false$_{\mathcal{R}}$ at $(f, w)$, while $\Diamond(\Diamond p)$ will be true$_{\mathcal{R}}$ at $(f, w)$. There is no function $g$ which replicates this pattern in $\mathcal{D}$—i.e. which has $\Diamond p$ false$_{\mathcal{D}}$ at $g((f, w))$ and has $\Diamond(\Diamond p)$ true$_{\mathcal{R}}$ at $g((f, w))$. This is for the simple reason that it is a theorem of $\mathcal{D}$ that $\Diamond \varphi \leftrightarrow \Diamond(\Diamond \varphi)$, since $\Diamond(\Diamond \varphi)$ is true$_{\mathcal{D}}$ at $(s, x)$, for any $x$, iff $\Diamond \varphi$ is true$_{\mathcal{D}}$ at $(s, w')$ for some $w' \in s$ iff $\varphi$ is true$_{\mathcal{D}}$ at $(s, w'')$ for some $w'' \in s$ iff $\Diamond \varphi$ is true$_{\mathcal{D}}$ at $(s, x)$ for any $x$. Thus there is no function from the indices of $\mathcal{R}$ to those of $\mathcal{D}$ which preserves truth for all $\varphi \in L^0$, and thus by Fact 4.3.1 we have $\mathcal{R} \not\preceq_{L^0} \mathcal{D}$.

Fact 4.3.7. $\mathcal{R} \not\preceq_{L^0} \mathcal{D}$.

Proof. We assume that any two worlds differ on the truth of some atomic sentence. Now consider three different modal bases $f, f'$, and $f''$, with $f(w) = f'(w) = f''(w) = \emptyset$. Consider any function $g$ from relational indices to domain indices which preserves truth for $\varphi \in L^0$. Suppose that $g((f, w)) = (s, x)$, $g((f', w)) = (s', x')$, and $g((f'', w)) = (s'', x'')$, with $(s, x), (s', x')$, and $(s'', x'')$ all different. Since all worlds differ on the truth of some atom, we know that $x = x' = x'' = w$, else we would have that at least one of $g((f, w)), g((f', w))$, or $g((f'', w))$ differs from its preimage on the truth of some atom, contrary to the assumption that $g$ preserves truth. So we must have that $s \neq s'$ and $s \neq s''$ and $s' \neq s''$. It is easy to see that, for any atom $p$, $\Diamond p$ is false$_{\mathcal{R}}$ at all of $(f, w), (f', w),$ and $(f'', w)$. But there are only two $\mathcal{D}$-indices with $w$ as their world parameter which make $\Diamond p$ false$_{\mathcal{D}}$ for every atom $p$, namely $(\emptyset, w)$ and $(\{w^f\}, w)$, where $w^f$ is the world such that for every atomic sentence $p$, $v(p, w^f) = 0$. And so either $(s, x), (s', x')$, or $(s'', x'')$ will make $\Diamond p$ true$_{\mathcal{D}}$ for some $p$, contrary to the assumption that $g$ preserves truth. Thus any truth-preserving function must take two of $(f, w), (f', w),$ and $(f'', w)$ to the same $\mathcal{D}$ index, and thus will fail to be an injection. Then Fact 4.3.7 follows by Fact 4.3.1.6

Fact 4.3.8. $\mathcal{U} \preceq_{L^0} \mathcal{R}$.

Proof. We assume that $\mathcal{R}$ and $\mathcal{U}$ have the same valuation functions, and the same stock of worlds, except for two in $\mathcal{R}$ not in $\mathcal{U}$, which we call $\alpha$ and $\beta$ (and which can receive any atomic valuation). We assume that, for any set of atomic sentences, there is exactly one world in $\mathcal{U}$ which verifies exactly those sentences. Let $w^f$ denote the world in $\mathcal{U}$ which is such that, for every atomic sentence $p$, $v(p, w^f) = 1$, and let $w^f$ denote the world in $\mathcal{U}$ which is such that, for every atomic sentence $p$,

---

6We could try to avoid this by adding copies of worlds to the stock of worlds in $\mathcal{D}$, but the result would be unattractive: we would have to hugely multiply worlds in $\mathcal{D}$ to avoid this result, in a way which looks completely unmotivated. Similar points go for a number of results below.
\[ v(p, w^f) = 0. \] Where \( \Phi \) is a set of atomic sentences, we also use \( \Phi \) to refer to the unique world from \( \mathcal{U} \) which verifies those sentences. For any sets of worlds \( r, s, c \), let \( f^r_{(s,c)} \) be the function which takes every world from \( \mathcal{U} \) to \( r \), and which takes \( \alpha \) to \( s \) and \( \beta \) to \( c \). We can then define a witness function \( g \) as follows: for any pair \( (s, c) \) of contexts, with \( p \) ranging over atomic sentences:

\[
g((s, c)) = \begin{cases} 
  \{ f^s_{(s,c)}, \{ p : \forall w' \in s : p \in w' \} \} & \text{iff } s = c \neq \emptyset \\
  \{ f^c_{(s,c)}, \{ p : (\forall w' \in c : p \in w') \land (\forall w'' \in (s \setminus c) : p \notin w'') \} \} & \text{iff } c \subseteq s \land c \neq \emptyset \\
  \{ f\{w:(w\neq\alpha)\land(w\neq\beta)\land(\forall p\in w\land\forall w'\in \emptyset : p \notin \emptyset)\}, \{ p : \forall w' \in s : p \notin w' \} \} & \text{iff } s \neq c \land c = \emptyset \\
  \{ f^w_{(s,c)}, w^f \} & \text{iff } s = c = \emptyset 
\end{cases}
\]

First note that \( g \) is an injection, since each pair of contexts is taken to an index whose modal base is tagged with that pair. \(^7\)

Now note that, for any sentence \( \varphi \in \mathcal{L}^\emptyset \) and \( \mathcal{U} \)-index \( i \), \( \varphi \) is true_\( \mathcal{U} \) at \( i \) iff \( \varphi \) is true_\( \mathcal{R} \) at \( g(i) \). To see this, consider first atomic \( q \). Atomic \( q \) is true_\( \mathcal{U} \) at \( (s, c) \) iff \( c \) is the result of removing all and only \( q \)-worlds from \( s \). \(^8\)

- if \( s = c \neq \emptyset \), then this holds iff all the worlds in \( s \) are \( q \)-worlds iff \( q \in \{ p : \forall w' \in s : p \in w' \} \) iff \( q \) is true_\( \mathcal{R} \) at \( g((s, c)) = \langle f^s_{(s,c)}, \{ p : \forall w' \in s : p \in w' \} \rangle \);
- if \( c \subseteq s \land c \neq \emptyset \), then this holds iff all the worlds in \( c \), but none of the worlds in \( s \setminus c \), are \( q \)-worlds, iff \( q \in \{ p : (\forall w' \in c : p \in w') \land (\forall w'' \in (s \setminus c) : p \notin w'') \} \), iff \( q \) is true_\( \mathcal{R} \) at \( g((s, c)) = \langle f^c_{(s,c)}, \{ p : (\forall w' \in c : p \in w') \land (\forall w'' \in s : p \notin w'') \} \rangle \);
- if \( c \notin s \), then this never holds, in which case \( q \) is also false_\( \mathcal{R} \) at \( g((s, c)) = \langle f^w_{(s,c)}, \{ p : \forall w' \in s : p \notin w' \} \rangle \);
- if \( c = \emptyset \land s \neq c \), then this holds iff no world in \( s \) is a \( q \)-world iff \( q \in \{ p : \forall w' \in s : p \notin w' \} \) iff \( q \) is true_\( \mathcal{R} \) at \( g((s, c)) = \langle f^w_{(s,c)}, \{ p : \forall w' \in s : p \notin w' \} \rangle \);
- if \( s = c = \emptyset \), this holds in any case whatsoever, in which case \( q \) is also true_\( \mathcal{R} \) at \( g((s, c)) = \langle f^w_{(s,c)}, \{ p : \forall w' \in s : p \notin w' \} \rangle \).

Consider next sentences of the form \( \Diamond q \), for atomic \( q \). In the update semantics, \( \Diamond q \) is treated as a "test": it takes a context \( c \) and returns \( c \) unchanged just in case \([ q ]_\mathcal{U}(c) \neq \emptyset \), and otherwise returns \( \emptyset \). That means that, for atomic \( q \), \( \langle s, c \rangle \in \{ \Diamond q \}_\mathcal{U} \) iff

\(^7\)It may be possible to construct an injective truth-preserving function without this convenience, but the construction would be substantially more complicated.

\(^8\)For atomic \( p \), where \( v \) is our valuation function, a \( p \)-world is a world \( w \) where \( v(p, w) = 1 \); a \( \beta \)-world is a world where \( v(p, w) = 0 \).
• (i) there is a q-world in \(s\) and \(s = c\); then \(\Diamond q\) is true\(_R\) at \(g((s, c)) = \langle f_{(s, c)}^q, \{p : \forall w' \in s : p \in w'\}\rangle\), since \(f_{(s, c)}^q(w)\) will contain a q-world, for any \(w\) other than \(\alpha\) or \(\beta\), and since the world parameter of \(R\)-indices in the image of \(g\) are never equal to \(\alpha\) or \(\beta\), by construction;

• or (ii) there is no q-world in \(s\), and \(c = \emptyset\).

Suppose first that \(s = c\). Then \(\Diamond q\) will be true\(_R\) at \(g((s, c)) = \langle f_{(s, c)}^q, \{p : \forall w' \in s : p \in w'\}\rangle\), since the fact that \(q\) is false throughout \(s\) ensures that \(q\) will be true at some world in \(\{w : (w \neq \alpha) \land (w \neq \beta)\} \land (\forall p : p \in w \rightarrow \forall w' \in s : p \notin w')\).

Suppose next that \(s = \emptyset\). Then \(\Diamond q\) will be true\(_R\) at \(g((s, c)) = \langle f_{(s, c)}^q, w^f\rangle\), since \(v(q, w) = 1\).

Next, suppose that \(\Diamond q\) is false\(_U\) at \((s, c)\); this will hold iff:

• \(s\) doesn’t contain a q-world and \(c \neq \emptyset\); then either \(g((s, c)) = \langle f_{(s, c)}^q, \{p : \forall w' \in s : p \in w'\}\rangle\) or \(g((s, c)) = \langle f_{(s, c)}^q, \{p : (\forall w' \in c : p \in w') \land (\forall w'' \in s : p \notin w'')\}\rangle\) or \(g((s, c)) = \langle f_{(s, c)}^q, w^f\rangle\); in any case, since \(v(q, w) = 0\) at every \(w\) in \(s\) and (trivially) at every \(w\) in \(\emptyset\), \(\Diamond q\) will be false\(_R\) at \(g((s, c))\); or

• \(s\) contains a q-world and \(s \neq \emptyset\); then either \(g((s, c)) = \langle f_{(s, c)}^q, \{p : (\forall w' \in c : p \in w') \land (\forall w'' \in s : p \notin w'')\}\rangle\) or \(g((s, c)) = \langle f_{(s, c)}^q, w^f\rangle\) or \(g((s, c)) = \langle f_{(s, c)}^q, w^f\rangle\). If the first or second, then \(\Diamond q\) will be false\(_R\) at \(g((s, c))\), since \(\emptyset\) doesn’t contain a q-world. If the third, then \(\Diamond q\) will be false\(_R\) at \(g((s, c))\), since \(\{w : (w \neq \alpha) \land (w \neq \beta)\} \land (\forall p : p \in w \rightarrow \forall w' \in s : p \notin w')\) will not contain a q-world, since there is a q-world in \(s\).

Finally, it is a theorem of the update semantics that, for any \(\varphi \in L^U\), \(\Diamond (\Diamond \varphi) \leftrightarrow \Diamond \varphi\). And this will also be true\(_R\) relative to any point in the image of \(g\), since for any such point \(\langle f, w \rangle\), \(f\) is constant within \(f(w)\), i.e. \(\forall w' \in f(w) : f(w') = f(w)\); this follows since \(f(w)\) will never contain either \(\alpha\) or \(\beta\), by construction, and \(f\) is always a constant function when on the domain of worlds which excludes \(\alpha\) and \(\beta\). From this it follows that \(\Diamond (\Diamond \varphi) \leftrightarrow \Diamond \varphi\) will be true at any point in the image of \(g\). And thus we can conclude that \(\Diamond \varphi\) is true\(_U\) at \(i\) iff \(\Diamond \varphi\) is true\(_R\) at \(g(i)\), for any \(\varphi \in L^U\).

Thus \(g\) is an injection from the indices of \(U\) to those of \(R\) which preserves truth for all sentences of \(L^U\), and thus by Fact 4.3.1 we have \(U \succeq L^U R\). \(\Box\)

**Fact 4.3.11.** \(U \not\leq L^U D\).

\(\text{Note that not every operator which can be added to} \ U \ \text{will be well-defined if we want the intension of any sentence in} \ U \ \text{to be a function from contexts to contexts, rather than just a relation; there are different approaches within broadly update-style frameworks to this question (e.g. Heim 1983 vs. Groenendijk and Stokhof 1991).} \)
Proof: Consider the three update indices \( \langle \emptyset, \{ w \} \rangle, \langle \emptyset, \{ w' \} \rangle, \langle \emptyset, \{ w'' \} \rangle \), with \( w, w', w'' \) all different. It is easy to see that these three indices all make \( p \) false\(_U\), for any atomic \( p \in \mathcal{L}^0 \); they also make \( \Diamond \varphi \) false\(_U\), for any \( \varphi \in \mathcal{L}^0 \), and thus make every sentence in \( \mathcal{L}^0 \) false\(_U\). There are, however, only two indices in \( \mathcal{D} \) which make every sentence in \( \mathcal{L}^0 \) false\(_D\), namely \( \langle \emptyset, w^f \rangle \) and \( \langle \{ w', w^f \}, w^f \rangle \), where \( w^f \) is, again, the world where, for every atomic sentence \( p \), \( v(p, w^f) = 0 \) (we continue to assume that any two worlds differ on some atom). Thus any truth-preserving function will have to map at least two of the \( \mathcal{U} \)-indices in question to the same \( \mathcal{D} \)-index, and thus will fail to be an injection. Thus Fact 4.3.11 follows by Fact 4.3.1.

\( \Box \)

Fact 4.3.12. \( \mathcal{D} \not\leq \mathcal{L}^0 \mathcal{U} \).

Proof. There are exactly three \( \mathcal{U} \)-indices which make all sentences in \( \mathcal{L}^0 \) true\(_U\), namely \( \langle \emptyset, \emptyset \rangle \), \( \langle \{ w^f \}, \emptyset \rangle \), and \( \langle \{ w^f \}, \{ w^f \} \rangle \). Assuming that there is at least one possible world in our stock of worlds other than \( w^f \) and \( w^t \), there are more than three \( \mathcal{D} \)-indices which make all sentences in \( \mathcal{L}^0 \) true\(_D\): these include \( \langle W, w^f \rangle \) and \( \langle \{ w^f \}, w^f \rangle \), as well as \( \langle s, w^f \rangle \) for any \( s \) such that \( \{ w^f \} \subseteq s \subseteq W \). Thus any function from the indices of \( \mathcal{D} \) to those of \( \mathcal{U} \) which preserves truth for all sentences in \( \mathcal{L}^0 \) will have to map more than three \( \mathcal{D} \)-indices to three \( \mathcal{U} \)-indices, and so will fail to be an injection, and thus, by Fact 4.3.1, \( \mathcal{D} \not\leq \mathcal{L}^0 \mathcal{U} \).

\( \Box \)

Fact 4.3.14. \( \mathcal{S} \prec \mathcal{L}^0 \mathcal{D} \).

Proof. Let \( \mathcal{S} \) have the same valuation function and stock of worlds as \( \mathcal{D} \), and continue to assume that each set of atomic sentences from \( \mathcal{L}^0 \) corresponds to exactly one world which verifies just those sentences, and which we treat for present purposes just as that set of sentences. Let the function \( g \) take any information state \( s \) to \( \langle s, \bigcap s \rangle \). For atomic \( p \), \( p \) is true\(_S\) at \( s \) iff for all \( w' \in s \), \( v(p, w') = 1 \), iff \( p \in \bigcap s \), iff \( p \) is true\(_D\) at \( g(s) = \langle s, \bigcap s \rangle \). For atomic \( p \), \( \Diamond p \) is true\(_S\) at \( s \) iff \( s \) contains a \( p \)-world iff \( \Diamond p \) is true\(_D\) at \( g(s) = \langle s, \bigcap s \rangle \). Finally, it is a theorem of both \( \mathcal{S} \) and \( \mathcal{D} \) that \( \Diamond (\Diamond \varphi) \leftrightarrow \Diamond \varphi \), and so we know that for any \( \varphi \in \mathcal{L}^0 \), \( \Diamond \varphi \) will be true\(_S\) at \( i \) iff \( \Diamond \varphi \) is true\(_D\) at \( g(i) \). Note finally that \( g \) is an injection: for any \( s \) and \( s' \), if \( s \neq s' \), then the first elements of \( g(s) \) and \( g(s') \) will differ. Thus by Fact 4.3.1 we have \( \mathcal{S} \preceq \mathcal{L}^0 \mathcal{D} \).

But we do not have the converse: \( \mathcal{D} \not\leq \mathcal{L}^0 \mathcal{S} \). Consider any function from the indices of \( \mathcal{D} \) to the indices of \( \mathcal{S} \). Consider three indices \( \langle \emptyset, w \rangle \) and \( \langle \emptyset, w' \rangle \), and \( \langle \emptyset, w'' \rangle \) with \( w, w', w'' \) all distinct. For any \( \varphi \in \mathcal{L}^0 \), \( \Diamond \varphi \) is false\(_D\) at all these indices. The only indices which make \( \Diamond \varphi \) false\(_S\) for every \( \varphi \in \mathcal{L}^0 \) are \( \emptyset \) and \( \{ w^f \} \), and thus any truth-preserving function will have to take two of the \( \mathcal{D} \)-indices to the same \( \mathcal{S} \)-index, and thus will fail to be an injection; thus by Fact 4.3.1, \( \mathcal{D} \not\leq \mathcal{L}^0 \mathcal{S} \).

\( \Box \)

Fact 4.3.15. For any language \( \mathcal{L} \), \( \preceq \mathcal{L} \) is a partial pre-order.

194
Proof. \( \preceq_L \) will be transitive: if \( \mathcal{M} \preceq_L \mathcal{M}' \), witnessed by \( g \), and \( \mathcal{M}' \preceq_L \mathcal{M}'' \), witnessed by \( f \), then the composition of \( g \) with the limitation of \( f \) to the image of \( g \) will witness \( \mathcal{M} \preceq_L \mathcal{M}'' \). \( \preceq_L \) is reflexive, witnessed by the identity function. It is not anti-symmetric, since it is easy to see that there are different models \( \mathcal{M} \) and \( \mathcal{M}' \) such that \( \mathcal{M} \preceq_L \mathcal{M}' \) and \( \mathcal{M}' \preceq_L \mathcal{M} \) (for instance, two standard models for a language just comprising atomic sentences, with the same set of possible worlds but different valuations, may have this property with respect to atomic sentences). And it is not connected, since, as we saw in the comparison of \( D \) to \( U \), there are models which are incommensurable with respect to a given language.

Fact 4.3.17. \( D^3 \preceq_L^\circ \mathcal{R}^3 \).

Proof. We assume that the domains, set of worlds, and valuation functions for \( D^3 \) and \( \mathcal{R}^3 \) are the same. For any \( D^3 \)-index \( \langle a, s, w \rangle \), with \( a \) a variable assignment, \( s \) a set of worlds, and \( w \) a world, let \( g \) take \( \langle a, s, w \rangle \) to the \( \mathcal{R}^3 \)-index \( \langle a, f^s, w \rangle \), where \( f^s \) is again the constant function to \( s \). \( g \) will clearly witness \( D^3 \preceq_L^\circ \mathcal{R}^3 \); the proof is a straightforward generalization of the parallel result in the proof of Fact 4.3.2. But there is no truth-preserving injection in the other direction, for the same reasons given in the proof of Fact 4.3.4. Note moreover that the proof of Fact 4.3.7 can be extended to show that \( \mathcal{R}^3 \not\preceq_L^\circ D^3 \).

Fact 4.3.18. \( U^3 \preceq_L^\circ \mathcal{R}^3 \).

Proof. The proof that \( U^3 \preceq_L^\circ \mathcal{R}^3 \) goes by way of constructing a witness function \( g \) as follows:

\[
g(\langle a, (s, c) \rangle) = \begin{cases} 
\langle a, f^s_{(s,c)}, w \rangle : \forall R : v(R, w) = \bigcap \{ v(R, w') : w' \in s \} & \text{iff } s = c \neq \emptyset \\
\langle a, f^s_{(s,c)}, w \rangle : \forall R : v(R, w) = \bigcap \{ v(R, w') : w' \in (s \setminus c) \} & \text{iff } c \subseteq s \text{ and } c \neq \emptyset \\
\langle a, f^c_{(s,c)}, w \rangle : \forall R : v(R, w) = \bigcap \{ v(R, w') : w' \in (s \setminus c) \} & \text{iff } c \nsubseteq s \\
\langle a, f^c_{(s,c)}, w \rangle : \forall \forall R^n : v(R^n, w) = D^n \setminus \bigcup \{ v(R^n, w') : w' \in s \} & \text{iff } s \neq c \text{ and } c = \emptyset \\
\langle a, f^c_{(s,c)}, w \rangle : v(R^n, w) = D^n \setminus \bigcup \{ v(R^n, w') : w' \in s \} & \text{iff } s \neq c \text{ and } c = \emptyset 
\end{cases}
\]

\( R \) ranges over relation symbols in the vocabulary of \( L^\circ \); \( w_t \) is the world such that \( v(R^n, w_t) \) is the universal \( n \)-ary relation, for any \( n \)-place relation symbol \( R^n \); and \( w_f \) the world such that \( v(R^n, w_f) \) is the empty relation, for any \( R^n \). \( D \) is the domain of individuals (which, again, we assume to be the same in both models); \( \bar{d} \) ranges over ordered sequences of elements of \( D \); and \( \nu \) ranges over worlds.
not including α or β. \( f^a_{r,c} \) is defined as in the proof of Fact 4.3.8 (and thus we assume, again, that \( \mathcal{U}^3 \) has two worlds not in \( \mathcal{U}^3 \), which can be valued in any way; and that otherwise worlds correspond exactly to possible combinations of relations).

The proof that \( g \) is a witness function of \( \mathcal{U}^3 \subset \mathcal{L}_g^3 \) \( \mathcal{R}^3 \) goes very much as the proof of Fact 4.3.8 above. First note that \( g \) is an injection, since each pair of a variable assignment and a pair of contexts is taken to an index which preserves that variable assignment, and whose modal base is tagged with that pair of contexts. Next note that, for any atomic sentence with the form \( \exists S(s) \) \( \mathcal{C}V \), for any relation symbol \( S \), and \( \mathcal{U}^3 \)-index \( \langle a, \langle s, c \rangle \rangle \), \( S(\vec{x}) \) is true \( \mathcal{U}^3 \) at \( g(\langle a, \langle s, c \rangle \rangle) \). To see this, note that atomic \( \exists S(s) \) is true \( \mathcal{U}^3 \) at \( g(\langle a, \langle s, c \rangle \rangle) \) if \( c \) is the result of removing all and only worlds \( w \) from \( s \) where \( a(\vec{x}) \notin v(S, w) \) (where \( a(\vec{x}) = \langle a(x_1), a(x_2) \ldots a(x_n) \rangle \) when \( \vec{x} = \langle x_1, x_2, \ldots x_n \rangle \);

- if \( s = c \neq \emptyset \), then this holds iff, for all the worlds \( w \) in \( s \), \( a(\vec{x}) \in v(S, w) \) iff \( a(\vec{x}) \in v(S, w') \) \( \forall R : v(R, w) = \bigcap \{v(R, w') : w' \in s\} \) iff \( S(\vec{x}) \) is true \( \mathcal{U}^3 \) at \( g(\langle a, \langle s, c \rangle \rangle) = \langle a, f^a \rangle \), \( \forall w : v(R, w) = \bigcap \{v(R, w') : w' \in s\} \);

- if \( c \subseteq s \), then this never holds, in which case \( S(\vec{x}) \) is also false \( \mathcal{U}^3 \) at \( g(\langle a, \langle s, c \rangle \rangle) = \langle a, f^a \rangle \);
Suppose first that $s \neq c$. Then $\diamond S(\bar{x})$ will be true in $R$ at $g((a, \langle s, c \rangle)) = \langle a, f_{(s, c)}^{w}, w \rangle$, since the fact that $a(\bar{x}) \notin v(S, w)$, for every $w \in s$, ensures that for some world $w' \in \{w : (w \neq \alpha) \land (w \neq \beta) \land v(R, w) \rightarrow v(R, w') \}$, we have $a(\bar{x}) \in v(S, w')$.

- Suppose next that $s = c$. Then $\diamond S(\bar{x})$ will be true in $R$ at $g((a, \langle s, c \rangle)) = \langle a, f_{(s, c)}^{w}, w \rangle$.

Next, suppose that $\diamond S(\bar{x})$ is false in $R$ at $g((a, \langle s, c \rangle))$; this will hold iff:

- For every $w \in s$, $a(\bar{x}) \notin v(S, w)$ and $c \neq \emptyset$; then either $g((a, \langle s, c \rangle)) = \langle a, f^{w}_{(s, c)} \rangle$, $w : \forall R : v(R, w) = \cap \{v(R, w') : w' \in s\}$ or $g((a, \langle s, c \rangle)) = \langle a, f^{c}_{(s, c)} \rangle$, $w : \forall R : v(R, w) = \cap \{v(R, w') : w' \in c\} \cup \{v(R, w') : w' \in (s \setminus c)\}$.

In any case, since we know that $a(\bar{x}) \notin v(S, w)$ at every $w$ in $s$, and (trivially) at every $w$ in $\emptyset$, then $\diamond S(\bar{x})$ will be false in $R$ at $g((a, \langle s, c \rangle))$.

- or for some $w \in s$, $a(\bar{x}) \in v(S, w)$ and $s \neq c$; then either $g((a, \langle s, c \rangle)) = \langle a, f^{c}_{(s, c)} \rangle$, $w : \forall R : v(R, w) = \cap \{v(R, w') : w' \in s\} \cup \{v(R, w') : w' \in (s \setminus c)\}$ or $g((a, \langle s, c \rangle)) = \langle a, f^{w}_{(s, c)} \rangle$, $w : \forall R : v(R, w) = \cap \{v(R, w') : w' \in s\} \cup \{v(R, w') : w' \in (s \setminus c)\}$.

If the first or second, then $\diamond S(\bar{x})$ will be false in $R$ at $g((a, \langle s, c \rangle))$, since $\emptyset$ does not contain a world $w$ such that $a(\bar{x}) \in v(S, w)$. If the third, then $\diamond S(\bar{x})$ will also be false in $R$ at $g((a, \langle s, c \rangle))$, since there is no $w'$ in $\{w : (w \neq \alpha) \land (w \neq \beta) \land v(R, w) \rightarrow v(R, w') \}$ such that $a(\bar{x}) \in v(S, w')$, since we know there is some $w'' \in s$ such that $a(\bar{x}) \in v(S, w'')$.

Finally, it is a theorem of the update semantics that, for any $\varphi \in L^{3}$, $\diamond (\diamond \varphi) \leftrightarrow \diamond \varphi$. And this will also be true in $R$ relative to any point in the image of $g$. And thus we can conclude that $\diamond \varphi$ is true in $R$ at $i$ iff $\diamond \varphi$ is true in $R$ at $g(i)$, for any $\varphi \in L^{3}$.

Thus $g$ is an injection from the indices of $U^{3}$ to the indices of $R^{3}$ which preserves truth for sentences of $L^{3}$, and thus we have $U^{3} \preceq_{L^{3}} R^{3}$.

The proof that $R^{3} \preceq_{L^{3}} U^{3}$ will be as for Fact 4.3.4. \qed
Bibliography


MacFarlane, J. (2004). In what sense (if any) is logic normative for thought? Presented at the Central Division APA.


203


