Title
The Dynamics of Disagreement and Contradiction

Author
Skeels, Patrick

Publication Date
2023

Peer reviewed|Thesis/dissertation

The Dynamics of Disagreement and Contradiction

By

Patrick Daniel Skeels

Dissertation

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

in the

Office of Graduate Studies

of the

University of California

Davis

Approved:

_____

Adam Sennet, Chair

_____

Rohan French

_____

Hanti Lin

Committee in Charge

2023

This dissertation would not have been possible in the absence of the overwhelming support from my friends and family as well as my colleagues at UC Davis and elsewhere. First thanks goes to my committee of Adam sennet, Hanti Lin, Rohan French, Masoud Jasbi, and Justin Khoo. Special thanks to Adam, Hanti, and Rohan for patiently helping me navigate protracted periods of confusion over the past several years. I also owe thanks to David Copp, Robert May, Elaine Landry, Jim Griesemer, and G.J. Mattey for useful conversations during my time at UC Davis. Thanks as well to my present and former colleagues Tim Houk, Jordan Bell, Da Fan, I-Sen Chen, Natasha Hadal, Elizabeth Rard, S. Kaan Tabakci, Khang Ton, Rachel Body, Ramiel Tamras, Harrison Smith-Jaoudi, Jerome Ramagosa, and Francisco Martinez-Aviña, as well as members of the LLEMMMa and L^3 reading groups. Special thanks to Tyrus Fisher and Brandon Hopkins for very lengthy discussions. From UC Davis linguistics, thanks are due to Casey Felton, Harvey Qiu, Skyler Reese, Thomas Hardy, and Jack Hawkins. Thanks to Aaron Bentley, Christopher Masciari, and Bas van Fraassen, as well as audiences at the ILLC, The University of Illinois Urbana-Champagne, UCLA, and The University of Kent. Unique thanks to my mother and father, who somehow maintained faith in my financially and professionally ridiculous endeavor to pursue a PhD in philosophy. Thanks, lastly, to my wife, Carolyn, and my daughter, Evelyn. The quiet patience and tireless effort of the former has allowed me to thrive inside and outside of academia, while the presence of the latter has forced me to achieve levels of focus and discipline I would have otherwise thought impossible. If there is any value to be found in the pages that follow, you deserve the credit.

For Caro and Ev

Contents

1 Modal Disagreements and Epistemic Contradictions **6**

1.1 Introduction 7

1.1.1 Epistemic Contradiction 8

1.1.2 Modal Disagreement 11

1.1.3 Parallels 15

1.2 Relativism 18

1.3 Bounded Semantics 22

1.4 Diagnosis 27

1.5 Dynamic Semantics 28

1.6 Dynamic Solutions 32

1.6.1 Incoherence First Approach 35

1.6.2 Asymmetric Approach 37

1.7 Trouble 39

2 Dynamic 'Might' and Correct Belief **42**

2.1 Preliminaries 47

2.1.1 Modal Disagreement 47

2.1.2 Test Semantics 48

2.2 Correctness Conditions 53

2.2.1 Formative Correctness vs. Content Correctness 53

2.2.2 Candidates for Content Correctness 57
2.3 Rethinking Indecision 73
2.4 Multiple Content Norms 77
2.4.1 Locative Correctness and Informational Correctness 78
2.4.2 Problems Solved 83
2.5 Concluding Remarks 89

3 Context, Contradiction, and Consistency 91
3.1 Dynamic Semantics 95
3.1.1 Overview 95
3.1.2 Formal Details 96
3.1.3 Selected Applications 99
3.2 Violations of Non-Contradiction 101
3.3 Context and Idealization 104
3.3.1 Context, Canonically 104
3.3.2 Dynamic Contexts 109
3.4 Context and Contradiction 111
3.5 What it Might Mean and What it Doesn't Mean 114

4 A Puzzle About Dynamic Belief 118
4.1 Formal Details 122
4.1.1 Dynamic Semantics 122
4.1.2 Doxastic Notions 125
4.2 A Puzzle About Dynamic Belief 128
4.3 Potential Strategies 131
4.3.1 Adjusting Indecision 131
4.3.2 Inversion Cases 133
4.3.3 Coherence 134

4.4 Minimal Consequence 135
4.5 Concluding Remarks 139

5 Conclusion **140**

i Proofs for Chapter 3 **141**

Introduction

Everybody knows that Aristotelian two-value logic is fucked.

-Philip K. Dick *Valis*

This dissertation concerns dynamic semantics. Dynamic semantics rejects the idea that meanings are truth-conditions and instead treats meanings as context change potentials (CCPs) that provide instructions for updating states of information. This deviation from the truth-conditional account of semantic meaning yields logics that feature numerous non-classical behaviors. The program was largely developed in the 80s and 90s, and attitudes about these behaviors were suitably rebellious for the time period. van Benthem (1996), for instance, says the following after listing a number of classical laws that dynamic semantics invalidates.

> ...[N]ew religions need not be defined by old dogmas which they accept or reject. Their point may be precisely that these old dogmas are too crude.

Here we find the same critical attitude towards classical logic expressed in the introductory quote, but its passive acrimony is replaced with a more optimistic and industrious sense of progress. Like many, I find this outlook attractive. The 00s brought increased interest in epistemic modality and with it, an opportunity to make good on the promise of progress. Puzzles involving modal disagreement and epistemic contradiction dominated the literature at the time, and competing views had to jump

through hoops in order to achieve results that seemingly fell out of the simplest versions of dynamic semantics. Dynamic semantics, as well as variations thereupon, situated themselves as genuine contenders in the literature. The problem, so far as I can tell, is that as the literature shifted from questions about the behavior of dynamic systems, to questions about whether and how they could be applied, many of the non-canonical behaviors of dynamic semantics were swept under the rug. Instead, conversations emphasized solutions to localized puzzles. The fact that many of these puzzles rested upon certain canonical assumptions that dynamic semantics explicitly rejects remained unappreciated. Other non-classical behaviors that yielded questionable predictions were at best, underdiscussed, and at worst unrecognized. When pressed, puzzling results were often embraced, but rarely explained.

Defenders of dynamic semantics typically assumed that (or at least wrote as if) more or less traditional epistemic and normative principles applied equally well in dynamic frameworks. Select, non-classical behaviors of dynamic semantics were emphasized when convenient, while less convenient results were rarely mentioned, and almost never explained. Gone was the rebellious attitude expressed in the quotes above, and a large part of the literature defending the dynamic program seemed to suggest that, despite the vast chasm between dynamic and truth-conditional semantics, the program could be comfortably situated within a dialectic that concerned problems based on largely classical and truth-conditional assumptions. Few proponents of the program strove to address details concerning how and why dynamic contents could be made to fit neatly with normative and epistemic principles designed exclusively for truth-conditional contents. It seemed, at least to some extent, that this was either taken for granted, or pushed to the side to be dealt with later. Meanwhile, critics lamented the non-classical behavior of dynamic semantics, and went as far as criticizing proponents as failing to recognize these behaviors (there are good reasons to believe that these allegations have some legitimacy). Other criticisms, however,

amounted to little more than complaints of non-classicality.

My assessment is that proper adjudication of these disputes and development of the program each require a great deal of semantic, pragmatic, normative, and epistemic discussion that is theoretically upstream from the narrow language puzzles popular in the literature. This dissertation can be seen as an attempt to begin such a discussion. In so doing, I bring various unique features of dynamic semantics to the fore, some novel, others not, and attempt to explain this behavior in the best ways available to dynamic semantics, or at least as best I can. As the title suggests, most of these issues concern, or are motivated by, disagreements and contradictions, broadly construed. I conclude, perhaps unsurprisingly, that the proponent of dynamic semantics is forced to not only reject central canonical semantic principles, but, in many cases, epistemic and normative ones as well. I then provide my own principles in order to fill the gaps. The primary insight is that if the dynamic program is to move forward in a philosophically rigorous manner, the conflict between dynamic semantics the aforementioned canonical principles must be appreciated, even if my alternative proposals are not. The dissertation is structured as follows:

Chapter 1 sets the stage by introducing dynamic semantics and popular competing semantics for epistemic modals. Theories are compared with respect to two motivating puzzles: *modal disagreement* and *epistemic contradiction*. I recognize that dynamic semantics offers a unified explanation to both puzzles, while competing theories do not, offering it a minor but, heretofore unrecognized, advantage in terms of theoretical elegance.

Chapter 2 poses a challenge unique to dynamic semantics and the way it addresses puzzles of modal disagreement. I argue that the operative normative notion of correctness proposed by dynamic theorists fails to motivate the disagreement challenge that they aim to solve. I argue that dynamic semantics will need a notion of correctness which motivates the puzzle, if solving the puzzle is to have any merit. I

consider extant proposals for correctness, and conclude that they are unsatisfactory. I then make a novel proposal for correctness within dynamic frameworks that deviates substantially from traditional approaches and demonstrate that it properly motivates the puzzle.

Chapter 3 considers the non-classical behavior of dynamic semantics, and focuses on failures of the law of Non-Contradiction. I observe that proponents of dynamic semantics have offered no explanation for this behavior, and some critics consider this result to be strong evidence against the tenability of the dynamic program. I defend and explain failures of Non-Contradiction by comparing dynamic semantics and classical, truth-conditional semantics in terms of their idealizing assumptions. I demonstrate that dynamic semantics rejects context fixity, an idealizing assumption which truth-conditional semantics adopts. I then argue that any semantics which rejects context fixity should, by the classical semanticist's own lights, violate non-contradiction under certain circumstances. I then demonstrate that dynamic semantics violates Non-Contradiction in all and only those circumstances. I then consider further indirect evidence in support of this result. I close by suggesting that discussion of idealizing assumptions, common in the sciences, is similarly crucial to fruitful discussion in natural language semantics.

Chapter 4 introduces a novel epistemic puzzle that does not occur in static frameworks. The puzzle involves contents which are decisive yet non-committal. Content is decisive with respect to an issue iff an agent who believes that content must be decided on the issue. Content that is non-committal is content that does not commit the believing agent to a particular decision, even if they are decided. Dynamic semantics allows for the unique case where contents can be both decisive yet non-committal, i.e. if an agent believes such content, they must be decided, but the content itself does not commit the believing agent to a particular choice on the issue. This results in an epistemic puzzle whereby undecided agents can update with

decisive, non-committal content, and after update, are decided. To solve the puzzle, I appeal to a long forgotten notion of entailment, introduced in Veltman (1996), and largely ignored thereafter, called minimal entailment. After demonstrating how and why minimal entailment allows the dynamic semanticist to solve the puzzle, I recommend that dynamic semanticists more fully embrace the extant practice of utilizing multiple entailment relations within the same semantic framework.

Chapter 1

Modal Disagreements and Epistemic Contradictions

Synopsis: This chapter is primarily a vehicle for exegesis as well as a means to get our protagonist center stage. Herein, I consider two puzzles that are especially prominent in the literature concerning epistemic modals: *modal disagreements* and *epistemic contradictions*. After introducing the puzzles, I recognize striking similarities between each. I then consider two prominent views in the literature: *relativism* and *bounded semantics*, each of which offers a solution to one of the above puzzles. I then observe that the nature of both relativism and bounded semantics ensures that each view is unable to leverage its solution to its preferred puzzle to solve the other. The consequence is that each view can, at best, provide disjoint solutions to both puzzles. Given the similarities between epistemic contradictions and modal disagreements, I suggest that this is somewhat unsatisfying. I subsequently introduce the the primary subject of inquiry for this dissertation: *dynamic semantics*, and consider its capacity to solve the puzzles. I then highlight the underappreciated fact that, unlike competing views, dynamic semantics provides a unified answer to both puzzles. I close by observing that, despite my arguments in favor of dynamic semantics, deeper

problems remain.

1.1 Introduction

The canonical Kratzerian account of natural language modality concerns, among others, expressions including modal auxiliaries like the following:

(1) It might be raining.

(2) Sam should apologize.

According to the canon, each of the above, as well as other modal expressions, have the following form:

(3) $Modal(B)\varphi$

Where $Modal$ is a quantifier which ranges over a set of possible worlds. Different modals are associated with different quantifiers, e.g. "might" and "possibly" are existential, while "ought", and "must" are universal. B is a modal base.[1] B places a restriction on the set of worlds in the domain of the quantifier. For instance, epistemic modal bases only include worlds compatible with what is known, while deontic modal bases only include worlds compatible with what is best, obligatory, etc. Accordingly, the modal base determines the flavor of the modal. The last argument, φ, is the formula within the scope of the modal, called the prejacent. The value for B can, in some cases, be supplied by explicit phrases headed with things like "in view of", (Kratzer, 1977, 340). Consider:

(4) In view of what I know, it might be raining.

(5) In view of what is best, Sam should apologize.

[1] B may also include an ordering source. Ordering sources are not especially relevant for my present concerns, so they shall be suppressed henceforth. For details on ordering sources, see Kratzer (1981).

In cases like (1) and (2), however, the value for B is not made explicit at the level of surface grammar. Instead, the value of B is taken to be contextually supplied, (Kratzer, 1977, 342). This helps explain why (1) is naturally read as epistemic, and (2) is not. It further explains how modalized phrases often have multiple, alternative readings. It is also where contextualism gets it name, since the value of B is determined by some context c in cases where B is not made explicit.

This dissertation focuses exclusively on modals that are epistemic in flavor. In such cases, the relevant notion of possibility is constrained, in one way or another, by what is known. One of contextualism's many strengths is its ability to provide a unified account of natural language modality. However, in limited cases concerning epistemic modals, contextualism allegedly makes faulty predictions. The literature on the semantics of epistemic modals has, for the past twenty or so years, focused heavily on objections to contextualism and has provided numerous semantic accounts of epistemic modals which strive to answer these challenges. Two objections, and their associated puzzles, have figured especially prominently in this literature. Call these puzzles *epistemic contradiction* and *modal disagreement*. Consider each in turn.

1.1.1 Epistemic Contradiction

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

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

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

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

(6)-(9) have many of the hallmarks of contradictions. Their truth-conditions, if any, are unclear, they are infelicitous when asserted across contexts. (6) and (7) have the form $\Diamond\neg\varphi \wedge \varphi$ while (8) and (9) have the form: $\Diamond\varphi \wedge \neg\varphi$ where $\Diamond$ is an epistemic possibility operator. Call sentences with one of these forms epistemic contradictions,

(Yalcin, 2007, 983-984). Proponents of the challenge argue that a satisfying account of the semantics and pragmatics of epistemic modals should be able to predict and explain the infelicity of epistemic contradictions.

We may be inclined to think that epistemic contradictions are Moore paradoxical. They look similar to many paradigm examples of Moore paradoxical sentences, like (10) below.

(10) # It's raining, and I don't know it.

Moore paradoxes are commonly taken to be paradoxes of assertion. This is to say that Moore paradoxical sentences cannot be felicitously asserted, but they can be felicitously supposed, or embedded under conditionals or attitude reports, (Yalcin, 2007, 984).

(11) If it's raining and I don't know it, then I'll get wet.

(12) Jim believes that it's raining and I don't know it.

(13) Suppose that it's raining and I don't know it.

Utterances with embedded Moore paradoxical sentences can be perfectly felicitous, suggesting that there isn't anything deviant about their content. Rather, their deviance is best explained pragmatically, as problems only arise from their assertion, (Yalcin, 2007, 984). Epistemic contradictions, however, do not behave this way.

(14) # If it is raining and it might not be raining, then you'll get wet.

(15) # Jim believes both that it is raining, and that it might not be raining.

(16) # Suppose that it is raining and it might not be raining.

When embedded, epistemic contradictions project their infelicity. This is strong evidence against the idea that they are Moore paradoxical, and further suggests that the best explanation ought to appeal to semantic features of these sentences, (Yalcin,

2007, 985-988). This is some evidence that epistemic contradictions are contradictions proper, and we should thus expect the conjuncts of epistemic contradictions to be inconsistent. However, with classical assumptions, this would commit us to the following:

$$\Diamond\varphi, \neg\varphi \vDash \bot$$

This option is blocked, however, as it commits us to:

$$\Diamond\varphi \vDash \varphi$$

which expresses the factivity of "might", (Yalcin, 2007, 988). This is an absolute nonstarter as it means that every time something might be the case, it is the case. Thus, our most immediate semantic and pragmatic explanations seem to fail, and the challenge is to predict and explain the infelicity of epistemic contradictions, both naked and embedded under conditionals, attitude verbs, and supposition, without incurring commitment to the factivity of "might".

The plot thickens when we consider arguments from Yalcin (2015). Here, Yalcin argues that speaker judgments suggest that sentences of the form $\varphi \wedge \Diamond\neg\varphi$ and $\neg\varphi \wedge \Diamond\varphi$ were infelicitous, while sentences of the form $\Diamond\varphi \wedge \neg\varphi$ and $\Diamond\neg\varphi \wedge \varphi$ were considered less so, (Yalcin, 2015, 498).[2] In brief:

$$\# \ \varphi \wedge \Diamond\neg\varphi$$

$$\# \ \neg\varphi \wedge \Diamond\varphi$$

$$?? \ \Diamond\varphi \wedge \neg\varphi$$

$$?? \ \Diamond\neg\varphi \wedge \varphi$$

[2]The data supporting these claims can be found in Knobe and Yalcin (2014). Similar intuitions motivate the test semantics for epistemic modals, developed in Veltman (1996). Here Veltman observes that claims like "Adam might be at the door... No, he isn't," are felicitously utterable, while the opposite ordering: "Adam isn't at the door... he might be," are comparatively infelicitous.

Epistemic contradictions where the epistemically modalized conjunct was first, were reported to be more felicitous than the opposite ordering. So, not only are epistemic contradictions infelicitous, but the order seems to matter as well. The puzzle is to explain the projected infelicity of epistemic contradictions without appealing to classical inconsistency. The graded infelicity of epistemic contradictions based on the ordering of the conjuncts demands prediction and explanation as well. Call this puzzle Epistemic Contradiction.

It is often taken that canonical contextualism fails to properly address epistemic contradictions since, according to contextualism, $\Diamond\varphi$ and $\neg\varphi$ are consistent, and changing the logic to treat them as inconsistent would yield the factivity of "might". Similarly, the symmetry of classical conjunction, often adopted by contextualists, disallows for any semantic discrepancies between $\Diamond\varphi \wedge \neg\varphi$ and $\neg\varphi \wedge \Diamond\varphi$. Accordingly, numerous alternative semantics have been provided in order to predict and explain the infelicity of epistemic contradictions.[3]

1.1.2 Modal Disagreement

It has also been argued that contextualism fails to predict certain cases of disagreement. Consider the following vignette.

Context: Holmes and Watson are trying to solve a murder. Holmes knows that Moriarty is not the murderer, but Watson is undecided.

(17) a. Watson: Moriarty might be the murderer.

b. Holmes: No, that's not right. Moriarty is not the murderer.

Numerous criticisms of the canonical contextualist semantics for epistemic modals rest on the claim that Holmes *disagrees* with Watson in this scenario. Call such cases

[3]See Yalcin (2007), Willer (2013), Mandelkern (2019), Moss (2015), and Moss (2018).

modal disagreements.[4] It is argued that we should consider the above to be a genuine disagreement for two primary reasons:

1. **Normative Assessment**: Holmes believes, and is licensed to assert, something like "No, that's wrong," directed at the content of Watson's assertion. More specifically, he judges the content that Moriarty might be the murderer as incorrect.[5] Crucially, this normative judgment is leveled at the epistemic modal claim in its entirety, and not merely the prejacent.[6]

2. **Simultaneous Belief**: Holmes cannot simultaneously believe the content of his own assertion, as well as Watson's. This is to say that he cannot rationally believe both that Moriarty might be the murderer, and that he isn't. Thus, Holmes' belief prevents him from rationally believing the content of Watson's assertion.

The above are taken to be hallmarks of disagreement, and can be found in uncontroversial cases of disagreement where some agent utters p and another agent utters $\neg p$. Such hallmark cases are often leveraged to support the claim that disagreement is grounded in inconsistency, and thus, instances of disagreement should be explained in terms of inconsistency. Accordingly, such critics place the onus for predicting and explaining disagreement squarely upon the semantics.

[4]Several variations of this scenario exist. The version I currently articulate is structurally identical to the one from MacFarlane (2011) which concerns spoken disagreements. However, not all versions of the puzzle work this way. See Egan et al. (2005) for a version that questions whether the contents of the beliefs are true. Lennertz (2019) focuses on the doxastic states of each agent.

[5]See MacFarlane (2011) and Lennertz (2019).

[6]There are several reasons to think that Holmes' normative assessment is directed at the content of Watson's assertion, and not his utterance thereof, or just the prejacent of Watson's claim. The first is that Holmes, can often say something like "What you said is wrong." Watson has not asserted the prejacent of the modal, but the full modalized phrase. Secondly, insofar as Watson does not know the identity of the murderer, he seems perfectly justified in believing and asserting that Moriarty might be the murderer. Indeed, Holmes could similarly assert something like "It's not the case that Moriarty might be the murderer, because he isn't. This, or similar utterances are the direct negation of the content of Watson's claim, are not mere rejections of the prejacent.

The above scenario counts as an objection to an older variant of Kratzerian contextualism where the contextually supplied value of a bare epistemic modal is just the speaker's information.[7] According to such a view, (17a) is equivalent to something like "For all I (Watson) know, Moriarty might be the murderer." or "I (Watson) don't know that Moriarty is not the murderer." Each of the previous are perfectly consistent with Moriarty not being the murderer. Similarly, Holmes is perfectly capable of simultaneously believing that Moriarty isn't the murderer, and for all Watson knows, he might be. Lastly, Watson seems perfectly justified in believing that it is possible, for all he knows, that Moriarty is the murderer, when he does not know that he isn't. This leaves Holmes' capacity to judge the content of Watson's assertion as incorrect unexplained. Thus, proponents of the challenge argue that this variant of contextualism fails to predict and explain the disagreement between Holmes and Watson.

The contextualist, however, is not restricted to the claim that the modal base is always and only determined by the speaker's information. Rather, it is often argued that the knowledge of other agents can be included in the modal base of epistemic "might." This strategy amounts to contextually widening the relevant epistemic community relative to which the modal is evaluated. The proponent of such a strategy can argue that (17a), in context, is more properly read as "For all we (Watson and Holmes) know, Moriarty might be the murderer." Should Holmes believe that Moriarty is not the murderer, then this would be inconsistent with Holmes' information, and the disagreement is predicted. Critics of contextualism have developed a variant of the puzzle that aims to challenge this strategy.

Context: Moriarty has faked his own death, and planted convincing evidence that he was killed. Moriarty has planted a bug near the evidence and is listening to the police conversation as Lestrade investigates the scene. Moriarty can hear the conversation

[7] MacFarlane (2011) calls this variant *solipsistic contextualism.*

of the investigators, but not vice versa.

(18) a. Lestrade: Moriarty might be dead.

b. Moriarty: Incorrect, Mr. Lestrade. I am very much alive.

It is argued that in this variant of the puzzle, Moriarty disagrees with Lestrade, and is able to do so for the same reasons that Holmes disagrees with Watson. Accordingly, we should expect the contents of Lestrade's utterance to be inconsistent with Moriarty's utterance/belief. However, unlike the previous case, it strikes as implausible that the modal base of (18a) can be expanded to include Moriarty. If it did, (18a) would mean something like "For all we (Lestrade, the other investigators *and Moriarty*) know, Moriarty might be dead." This interpretation not only strikes as odd (why would Lestrade include Moriarty as a member of the epistemically relevant community?) but unjustified. How could Lestrade include Moriarty's epistemic state if he does not know whether he is alive or dead? Thus, critics take this case to place the contextualist in the following dilemma:

1. Predict the disagreement, and treat (18a) and (18b) as inconsistent, at the cost of making Lestrade's putatively justified assertion unjustified.

2. Restrict the relevant epistemic community such that it does not include Moriarty, vindicating Lestrade's assertion at the cost of making the content of Lestrade's utterance perfectly consistent with Moriarty's, thus failing to predict the disagreement.

Critics of contextualism have argued that there is no way out of the dilemma for contextualists, and have used this to motivate various alternative semantics.

A temporal variant of the puzzle can also be generated.

Context: Watson is trying to solve the murder, and is pouring through the evidence.

(19) a. Watson: Moriarty might be the murderer.

b. * *Watson discovers evidence that conclusively absolves Moriarty of guilt.*

c. Watson: Okay, I was wrong. Moriarty isn't the murderer.

Upon uttering (19a) Watson entertains the possibility that Moriarty is the murderer. However, upon uncovering the evidence, he learns that Moriarty is not the murderer. This compels him to retract the claim in (19a), even though, for all he knew at the time, Moriarty might have been the murderer. Regardless of whether or not we are inclined to say that Watson disagrees with his previous self, we do seem obligated to say that Watson retracts his previous claim, and does so in a way that is similar, in all the relevant respects, to a disagreement, i.e. Watson is inclined to make a negative normative assessment of his previous belief, and cannot concurrently hold his previous belief and his new one. This, as before, is taken to be indicative of inconsistency, however, the contents of (19a) and (19c) are perfectly consistent according to the contextualist account and the retraction remains unexplained by both variants of contextualism. Call this puzzle Modal Disagreement. Should we take this objection seriously, then a satisfying semantics for epistemic modals should solve the puzzle by predicting and explaining instances of modal disagreement. Crucially, these explanations should also vindicate the two claims that motivate the disagreement, normative judgment and simultaneous belief.

1.1.3 Parallels

The similarities between the two puzzles are worth emphasizing. Each puzzle involves an epistemically modalized claim and the negation of the prejascent of the modal: $\Diamond\varphi$ and $\neg\varphi$.[8] In each case, there is a tension between the two formula. In the case of epistemic contradiction, both sentences embedded under conjunction yield a sentence whose behavior is remarkably similar to that of a contradiction. In the case

[8]Or, alternatively, they involve a claim, and the negation of that claim embedded under an epistemic modal: φ and $\Diamond\neg\varphi$.

of modal disagreement, when each sentence is thought/uttered by different speakers (or the same speaker at different times), there is a disagreement. Both observations drive us to the conclusion that $\Diamond\varphi$ and $\neg\varphi$ are jointly inconsistent, but this option is unavailable due to the factivity of "might". The parallels are significant enough to suggest that these are variants of the same puzzle.

Notice also the relationship between the felt infelicity of epistemic contradictions and the motivating claims for modal disagreement: normative assessment and simultaneous belief. The claim that one cannot simultaneously believe $\Diamond\varphi$ and $\neg\varphi$ strikes as intuitively related to the claim that $\neg\varphi \wedge \Diamond\varphi$ is contradictory or otherwise self defeating. This sense of self defeat is explained, in large part, by the observation that if $\neg\varphi$ is indeed, the case, then $\Diamond\varphi$ cannot be correct. Further observations about uncontroversial contradictions and disagreements suggest parallels between modal disagreements and epistemic contradictions as well. $\varphi \wedge \neg\varphi$ is taken to be the paradigm of contradiction.[9] Similarly paradigmatic cases of disagreement involve cases where one agent utters/believes φ and another utters/believes $\neg\varphi$. The contradictory nature of $\varphi \wedge \neg\varphi$ in addition to the disagreement that arises from φ and $\neg\varphi$ can be succinctly explained by the fact that φ and $\neg\varphi$ are inconsistent. The similarities between epistemic contradictions and modal disagreements, as well as their similarities to uncontroversial contradictions and disagreements suggest that a satisfying solution will provide a unified explanation to both puzzles, while also capturing the similarities between these cases and uncontroversial cases. This is to say that there is reason to prefer a unified explanation to the puzzles of modal disagreement and epistemic contradiction.

As I shall argue, many of the positions in the literature obviate the possibility of such a unified solution. In what follows, I consider a popular solution to the puzzle of modal disagreement as well as a compelling solution for epistemic contradictions.

[9]So long as φ expresses a proposition, even the highly non-classical dynamic theorist agrees. See Chapter 3 for details.

The first is relativism, defended most famously in MacFarlane (2011) and MacFarlane (2014). The views espoused here amount to a rejection of traditional contextualism about epistemic modals in favor of a semantics where epistemic modal claims are assessment sensitive. Upon this view, all of the above cases of disagreement and retraction are predicted. The second view considered can be called bounded semantics, which was introduced more recently in Mandelkern (2019). The bounded semanticist argues that epistemic contradictions are, in fact, classically inconsistent, while striving to avoid the factivity of "might". Bounded semantics can be seen as an attempt to vindicate the contextualist picture of epistemic modals, and more generally, the broadly Kratzerian framework of which this picture is a part. The suggestion is that the modal base of an epistemically modalized claim is bounded by the local context provided by the information contained in other sentences it is conjoined with. The result is that epistemic contradictions are classically inconsistent since the modal base of $\Diamond\varphi$ is constrained by the other conjunct $\neg\varphi$. As a consequence, $\Diamond\varphi \wedge \neg\varphi$ is a contradiction, and thereby inconsistent. The relevant result is that epistemic contradictions are inconsistent, and are thus predicted to be infelicitous, while "might" remains non-factive.

Despite the vast differences between relativism and bounded semantics, there is a similarity in their strategies. Each focuses on one of the two puzzles above, and localizes the tension between $\Diamond\varphi$ and $\neg\varphi$ in a way that solves that, and only that, puzzle. Relativism, as we shall see, claims that epistemic modals are assessment sensitive, and thus, the proposition expressed by $\Diamond\varphi$ can have distinct truth-values from different contexts of assessment. This allows the relativist to predict and explain instances of modal disagreement, spoken, eavesdropper, and otherwise. However, this strategy relies on the idea that $\Diamond\varphi$ be evaluated at different contexts of assessment (these contexts are informed by the different information of each speaker) in order to explain the felt tension that motivates the disagreement. Epistemic contradictions,

however, do not involve distinct contexts of assessment, which prevents the relativist from using the same strategy to predict the infelicity of epistemic contradictions. Bounded semantics, on the other hand, claims that $\Diamond\varphi$ and $\neg\varphi$ are jointly consistent, but when embedded under conjunction, the result is always a contradiction, and is thereby inconsistent. This allows bounded semantics to explain the infelicity of epistemic contradictions, even when they occur in extremely complex embedding environments. This explanation, however, fundamentally relies on conjunction, which modal disagreements need not involve. This prevents the bounded semanticist from using this explanation to predict and explain instances of modal disagreement. The concern is that, even if the views were modified to address the other problem, the other problem would have to be solved by some distinct mechanism. This would yield a disjoint solution to the problems. The next sections explore relativism and bounded semantics in greater detail.

1.2 Relativism

MacFarlane (2011) develops an *assessment sensitive*, or relativist, account of epistemic modals. The position is so called because it claims that the truth or falsity of an epistemic modal claim can be changed by evaluating it from different contexts of assessment, (MacFarlane, 2011, 160).[10] Relativism about epistemic modality is primarily motivated by the observation that contextualist accounts of epistemic modals are unable to predict and explain cases of modal disagreement.[11] Relativism is designed to do precisely this, and so the goal, at least with respect to epistemic modals, is to explain how speakers like Watson and Holmes disagree without being committed

[10]The view is further developed in MacFarlane (2014). A similar view is also proposed in Egan et al. (2005). In what follows I engage MacFarlane's version of relativism, but similar claims can be made about the other versions.

[11]See von Fintel and Gillies (2008) and Hawthorne (2007) for criticisms of the claim that modal disagreements are a genuine problem. Also see von Fintel and Gillies (2011) for a view which embraces the challenge, and strives to defend of contextualism against it. See Lennertz (2020) for a response.

to the factivity of "might".

The relativist argues that sentences of a language are evaluated relative to both contexts of utterance and contexts of assessment, (MacFarlane, 2011, 159-160). Contexts of utterance are nothing new, and are appealed to, in some respect, by just about every semantic framework.[12] Contexts of utterance, through some means or another, supply values for context sensitive fragments of a language. For instance:

(20) I'm hungry.

When uttered in context, (20) expresses the proposition that the speaker is hungry. What is important here is that the contextual value for "I" is fixed entirely by the context of utterance. Contexts of assessment, on the other hand, have to do with the context with respect to which a particular proposition is evaluated. More specifically, they determine whether a particular proposition is true or false relative to a context of assessment, (MacFarlane, 2011, 159, 164). Contexts of assessment are similar to contexts of utterance insofar as they are another parameter in the model with respect to which a sentence is interpreted. They are importantly different insofar as contexts of utterance determine what proposition is expressed by a sentence, while contexts of assessment play no role in determining what proposition is expressed. Instead, this parameter determines, with respect to some context of assessment, whether the proposition expressed by that sentence is true or false. This means that the very same proposition can vary in truth-value when evaluated at different contexts of assessment, and truth is thereby relative. For the relativist, only certain fragments of language trigger a shift the assessment parameter, (MacFarlane, 2011, 162). Accordingly, to defend relativism about epistemic modals is to defend the claim that epistemic modal sentences are assessment sensitive, and that epistemic modals themselves trigger such a parameter shift. This is to say that an epistemic possibility

[12]See Lewis (1980) and Kaplan (1989) for canonical ways in which this can be done. See Chapter 3 for a more detailed discussion.

claim can be true relative to some contexts of assessment, and false relative to others. Relativism, like contextualism, says that epistemic modal claims of the form $\Diamond\varphi$ are true just in case there is at least one world in the information state where φ is true, (MacFarlane, 2011, 164). Here an information state is the set of worlds consistent with some body of information. The relativist maneuver, and what separates them from the contextualist, is that the relativist does not rely on the context of use to determine the relevant information state. Instead, this is determined by the context of assessment. Thus, if an epistemic possibility claim is evaluated at a context of assessment where the prejacent is compatible with the information of the assessor, the proposition expressed by the claim will evaluate to true. However, the same proposition, when evaluated at a context of assessment where the prejacent is not compatible with the assessor's information, will evaluate to false at that context of assessment. Oversimplifying a bit, we can say that for the relativist, the modal base is supplied by the context of assessment rather than the context of utterance.

To see relativism in action, consider the conversation between Holmes and Watson. Watson's utterance of (17a) expresses the proposition that there is a world in the information state where "Moriarty is the murderer" is true. Since Watson is undecided, this condition holds in his information state. Thus, from Watson's context of assessment, (17a) is true. Holmes, however, believes that Moriarty is not the murderer and so there is no world in his information state where "Moriarty is the murderer" is true. Accordingly, the same proposition expressed by Watson's utterance of (17a) that is true relative to Watson's context of assessment, is false relative to Holmes'. In virtue of this, the relativist is able to predict that the two speakers disagree with respect to some particular proposition, (MacFarlane, 2011, 162). The same strategy can be applied to other cases as well. Cross-conversational and cross-temporal scenarios can also be predicted since the context of assessment parameter is in no way restricted by what the speaker or the speaker's expected audience knows and is always fixed by

the context of assessment. This allows the relativist to predict and explain cases of modal disagreement that the canonical contextualist semantics for epistemic modals cannot, (MacFarlane, 2011, 160). Importantly, this explanation vindicates the two observations that motivate the disagreement challenge. Because Holmes takes $\neg m$ to be true, $\Diamond m$ will always evaluate to false from Holmes' circumstance of evaluation. That these contents are false is sufficient to explain why Holmes judges this content to be incorrect.[13] This also neatly explains why Holmes cannot simultaneously believe the contents of Watson's assertion while maintaining his own. Any context of assessment that takes $\neg m$ to be true will ensure that $\Diamond m$ is false. Moreover, there are independent arguments for relativism with respect to other fragments of our language including taste predicates and future contingents, providing independent motivation for the move to relativism.[14]

Relativism performs quite well in its handling of the puzzle of modal disagreement. Unfortunately, the view falters with respect to epistemic contradictions. To see why, notice that relativism's ability to predict and explain cases of modal disagreement relies fundamentally on the fact that the context of assessment changes when the assessor changes. This is useful in cases that involve different speakers but is not helpful when the epistemic modal and the negation of its prejacent are embedded under conjunction in the same sentence, like (6)-(9). When the conflict exists in a single sentence, there isn't anything that is able to shift the context of assessment, and so relativism offers no solution to the puzzle of epistemic contradiction. The relativist must explain the infelicity of epistemic contradictions by some other means.

Here, it seems that the relativist's only option is to appeal to a pragmatic strategy. They can claim that to assert an epistemic contradiction, like (6)-(9) is to violate

[13]More explicitly, this can be explained by what can be called *Wedgewood's Thesis*: a belief is correct if and only if it is true, (Wedgewood, 2002, 267).

[14]See MacFarlane (2003), MacFarlane (2007), MacFarlane (2014) and Schaffer (2011) for alternative motivations for adopting relativistic semantics. My arguments here only apply to relativism about epistemic modals, and in no way target relativism at large, or relativism with respect to other domains.

some pragmatic principle. We might think that to assert the first conjunct of (6) one represents oneself as knowing that it is raining, while the second conjunct represents oneself as not knowing that it is raining, and that this violates some norm of conversation or rationality.[15] The good news is that the pragmatic principle that the relativist would need to adopt in order to explain the infelicity of epistemic contradictions is the very same that predicts the infelicity of Moore paradoxical assertions, and is, thereby, independently motivated. However, as Yalcin (2007) observes, such an explanation is only helpful for utterances of bare Moore sentences or epistemic contradictions. The rule does not explain the infelicity of embedded Moore sentences, as these are perfectly felicitous. Appeal to the same rule, unfortunately, also ensures that embedded epistemic contradictions will be predicted to be felicitous. With pragmatics alone, the relativist is unable to predict the infelicity of embedded epistemic contradictions, (Yalcin, 2007, pp.985). Importantly, the problem demands solutions to both bare and embedded cases, and with respect to the latter, relativism falls short. I suppose it is possible that the view be modified in such a way that it predicts the infelicity of epistemic contradictions as well. The crucial observation, however, is that this will have to be done through means that are distinct from the relativistic machinery developed thus far. This suggests that the solution to both puzzles will lack the unity I have earlier recommended.

1.3 Bounded Semantics

Bounded semantics is introduced in Mandelkern (2019) which aims to address puzzles involving epistemic contradictions. Mandelkern observes that many proposals which strive to predict the infelicity of epistemic contradictions fail to do so when two epis-

[15]MacFarlane (2011) claims that this explanation is available to a position called *solipsistic contextualism*, (MacFarlane, 2011, 145-146). MacFarlane presents relativism as a superior alternative to contextualist positions, solipsistic and otherwise, and while he does not cite this explanation in his direct defense of relativism, there is nothing stopping the relativist from appealing to this same explanation.

temic contradictions are embedded under disjunction. He further observes that the infelicity of these and other complex embeddings could be predicted should we simply embrace the claim that epistemic contradictions are genuine contradictions. No one, prior to Mandelkern, has embraced this claim, but Mandelkern bucks the trend, and argues that we ought to treat epistemic contradictions to be genuine contradictions. He does not argue, however, that $\Diamond\varphi$ and $\neg\varphi$ are inconsistent. Rather, he argues that $\Diamond\varphi$ and $\neg\varphi$ are inconsistent when embedded together under conjunction. This avoids the factivity of "might".

According to Mandelkern's proposal, the information contained in each conjunct of a conjunction is *bounded* by the information in the other conjunct. Epistemic modals are sensitive to this bounding, and the suggestion is that the modal base of an epistemic modal will include all of the information in the other conjunct. Thus, in the case of epistemic contradictions like $\Diamond p \wedge \neg p$ the modal base of $\Diamond p$ will contain the information $\neg p$, thus ensuring that the conjunction is a contradiction, (Mandelkern, 2019, 14). However, Mandelkern's strategy ensures that $\Diamond\varphi$ and $\neg\varphi$ remain perfectly consistent, but become inconsistent when embedded under conjunction. This avoids the factivity of "might" while predicting that epistemic contradictions are infelicitous, even under complex embeddings.

More formally, the bounded semanticists recommends the following semantic clause for epistemic "might":

(21) $[\![\text{might}_i\varphi]\!]^{g,k,w}$

- defined only if $\forall w' : g(i)(w') \subseteq k$;
- if defined, true iff $\exists w' \in g(i)(w) : [\![\varphi]\!]^{g,k,w'} = 1$

In the above, g is a variable assignment and k is a local context, (Mandelkern, 2019, 14). The bounded theorist, like the Kratzerian, treats epistemic 'might" as an existential quantifier over epistemically accessible worlds, and an epistemically modalized

sentence is true when at least one such world makes the prejacent true. The difference, is that the variable assignment that determines the modal base is only defined when it falls under a local context k. The role of this local context can be made clearer by looking at the clause for conjunction.

(22) $[\![\varphi \text{ and } \psi]\!] = 1$ iff $[\![\varphi]\!]^{g,k_g^{\psi},w} = 1$ and $[\![\psi]\!]^{g,k_g^{\varphi},w} = 1$

The clause for conjunction is identical to the traditional clause for conjunction, save that any material sensitive to a local context, e.g. an epistemic modal, is constrained by the information generated by the local context created by the other conjunct, (Mandelkern, 2019, 17). This ensures that the modal base of any epistemic modal claim embedded under conjunction will only be defined when the modal base includes the information contained in the other conjunct. This, in turn, ensures that for an epistemic contradiction $\Diamond\varphi \wedge \neg\varphi$, $[\![\Diamond\varphi]\!]$ is only defined at variable assignments that are constrained by the local context provided by $\neg\varphi$. However, any information state that supports $\neg\varphi$ will not have any accessible worlds where φ is true, and thus, whenever $[\![\Diamond\varphi \wedge \neg\varphi]\!]$ is defined, it is false. The same explanation applies to other orderings and variations as well. The end result is an account of epistemic modality that is Kratzerian in spirit, insofar as the semantics is relational, and epistemic modals maintain a contextually supplied modal base. This base however, can be bounded by information contained in any other conjuncts of a conjunction in which the modal is embedded. This feature sees inspiration from dynamic semantics, where epistemic modals are sensitive to updates in the same sentence prior to the modal. Unlike the dynamic theory, this sensitivity is symmetric and is not affected by the order of the conjuncts, (Mandelkern, 2019, 15). $\Diamond\varphi$ and $\neg\varphi$ can be perfectly consistent, so long as they are not brought together by conjunction. This allows for the bounded theorist to predict the infelicity of epistemic contradictions under several complex embeddings without incurring commitment to the factivity of "might".[16]

[16]This includes when epistemic contradictions are embedded under disjunction e.g., $(\Diamond\varphi \wedge \neg\varphi) \vee$

The heart of this strategy is not so much a shift in how we view the meanings of epistemic modals; it merely adds another way for them to be sensitive to context. The more substantial component of this strategy concerns a shift in the way we view conjunction. Namely, the insistence that each conjunct is sensitive to the information contained in the other. The result is a thorough account of epistemic modals which can be situated nicely within the broader Kratzerian framework, while avoiding the problems with epistemic contradiction that plague the traditional Kratzerian.

This strategy localizes the tension between $\Diamond\varphi$ and $\neg\varphi$ to cases involving the conjunction of the two, and $\Diamond\varphi$ and $\neg\varphi$ remain perfectly consistent. However, while the bounded theorist performs quite well with respect to epistemic contradictions, modal disagreements remain unexplained. In particular, eavesdropper cases like the one between Lestrade and Moriarty continue to pose problems. Bounded semantics only deviates from the Kratzerian in cases where the modal is embedded under conjunction. Eavesdropper cases need not involve conjunctions, and as a result, when Lestrade utters $\Diamond m$, as he does in (18a), the modal base is determined in traditional Kratzerian fashion. The modal base in (18a) very plausibly includes Lestrade's information, as well as the information of any of the other investigators with which he is conversing. By stipulation, it does not include anyone who knows $\neg m$, which rules out Moriarty. Nonetheless, Moriarty still appears to disagree with Lestrade despite the fact that the contents of (18a) and (18b) are perfectly consistent under the bounded theory. This leaves the disagreement unexplained.

Recall that the motivation for the claim that Moriarty disagrees with Lestrade comes from two separate claims. The first is that Moriarty cannot hold Lestrade's belief ($\Diamond m$) while simultaneously maintaining his own ($\neg m$). The second is that Moriarty is licensed to judge the content of Lestrade's belief to be incorrect. The bounded theorist is able to explain the the former, but not the latter. Notice that in virtue of

---

$(\Diamond\varphi \wedge \neg\varphi)$.

believing $\neg m$ Moriarty is unable to also, simultaneously believe $\Diamond m$. The bounded semanticist is able to explain why Moriarty rejects Lestrade's assertion because, if Moriarty were to accept it, he would plausibly be forced to believe $\neg m \wedge \Diamond m$, which, for the bounded theorist, is a contradiction. As Mandelkern puts it, "...conjoining two sentences can change the way that we interpret those sentences, (Mandelkern, 2019, 19). This is to say that the content of Lestrade's assertion of $\Diamond m$ will change when embedded under conjunction with Moriarty's other belief that $\neg m$, rendering Moriarty unable to adopt Lestrade's belief on pain of believing a contradiction. This allows for the bounded theorist to succinctly explain why Moriarty is unable to simultaneously maintain Lestrade's belief as well as his own. So much for the first motivation. The second, however, poses problems for the bounded theorist. While $\Diamond m$ is interpreted differently when embedded under conjunction, there are no conjunctions in Lestrade's assertion. Accordingly, the content of Lestrade's assertion is perfectly consistent with Moriarty's belief. Suppose, as is plausible for the contextualist (and thereby the bounded theorist), that the modal base of $\Diamond m$ includes Lestrade and other investigators with whom he is speaking. It follows on the contextualist account that the content of Lestrade's utterance can be captured by the claim that "In view of what we (Lestrade and the investigators) know, Moriarty might be the murderer." This, however, is perfectly consistent with Moriarty's belief, and it had better be, as this is the goal of Moriarty's ruse in the first place. Thus, it remains mysterious how Moriarty is able to judge the content of Lestrade's assertion to be wrong, when, upon recognizing Lestrade's ignorance, Moriarty himself, also believes Lestrade to be ignorant.

Crucially, the disagreement challenge relies fundamentally on the claim that Moriarty is able to judge the content of Lestrade's belief/assertion to be incorrect (normative assessment). Lestrade's assertion is more than just the prejacent of the modal, but does not include any conjunctions, and thereby, is not bounded. The bounded

theorist has no explanation as to why Moriarty is able to make this negative normative judgment. Failure to explain this normative judgment, even if Mandelkern explains the disagreement by some other means, does not answer the challenge.[17]

It is worth appreciating that this is not a knockdown argument against the bounded semanticist. There are a variety of means by which the disagreement may be plausibly explained that are available. Similarly, the semantics may be supplemented to explain Moriarty's judgement. What is less clear is that bounded semantics can be supplemented in such a way that provides a unified explanation for epistemic contradictions and modal disagreements. So long as the factivity of "might" is to be avoided, some modal disagreements will not be able to be explained in terms of inconsistency, and any solution available to the bounded theorist will, like the relativist, have to appeal to a distinct mechanism from the one the bounded theorist utilizes to explain epistemic contradictions.

1.4 Diagnosis

Both theories fared well with respect to the individual problem that they aimed to address. Relativism neatly explains modal disagreements, but sheds no light on epistemic contradictions. Bounded semantics was able to predict the infelicity of epistemic contradictions in a variety of embedding environments, but did little to explain modal disagreements and the observations that motivate them. Neither of these objections are necessarily damning, and each theory may have the capacity to be modified to better accommodate the data. However, I am not aware of any extant proposals in this vein.[18] My point, however, is that the strategy employed by each

[17]It is worth mentioning that MacFarlane's suggestion that interlocturs are inclined to judge epistemic modal claims with false prejacents as wrong or otherwise incorrect is backed up by empirical data in Khoo (2015).

[18]It is worth mentioning that, so far as I can tell, there is nothing stopping us from developing an amalgamation of both theories that we might call *bounded relativism*. Under such a theory, we could say that the information contained in one conjunct bounds admissible contexts of assessment

view to solve its preferred puzzle, prevents the very same strategy from applying to the other problem. Even if either view could be modified to better accommodate the data, it seems that a unified solution to both puzzles is a bridge too far, and the puzzles will need to be solved by appeal to distinct theoretical mechanisms.[19] I see no reason, however, to prefer distinct mechanisms for predicting such similar phenomena. This is to say that a unified account of the tension between $\Diamond\varphi$ and $\neg\varphi$ would not only better explain the data, but would enjoy greater theoretical elegance in the form of a unified explanation. Fortunately, we have such a theory at our disposal.

1.5 Dynamic Semantics

Mandelkern and MacFarlane go to great lengths to avoid the claim that $\Diamond\varphi$ and $\neg\varphi$ are inconsistent in order to similarly avoid the factivity of "might". The engineering choices employed by each strategy are largely motivated by principles of conservativity. Despite their more exotic features, both bounded semantics and relativism maintain that meaning is truth-conditional and that each view can be made to fit within the admittedly attractive Kratzerian framework. We need not, however, be so conservative. Should we adopt certain non-classical entailment relations, we are free to embrace the inconsistency of $\neg\varphi$ and $\Diamond\varphi$ without incurring the factivity of "might". Freedom from the shackles of conservatism allows us to more freely pursue our unified solution. Several theories implement precisely this strategy.[20] Dynamic semantics is one such theory and will be the primary subject of inquiry for this dissertation

Often called *update semantics*, or *test semantics*, dynamic semantics rejects the

of the other. This would, presumably, allow us to solve both puzzles in a satisfying and unified manner. The proposal strikes as unwieldy, runs the risk of being *ad hoc*, and seems to lack a degree of theoretical elegance. It may, nonetheless, have potential. As of now, no such view inhabits the literature, and I leave this possibility for future work.

[19] At least without tremendous revision to the original theory. I take the revision suggested in footnote above to be tremendous.

[20] Other accounts, distinct from the one I currently explore, pursue this strategy. They include, among others, Moss (2015). Moss (2018), and Yalcin (2007).

canonical claim that meaning is truth-conditional.[21] Introduced in Veltman (1996), and developed extensively thereafter, the test semantics has it that contents are not static propositions which are true or false. Instead, they are dynamic updates– functions from contexts to contexts–that provide instructions on how to update a body of information. Thus, meanings are not propositions but context change potentials (CCPs). Propositions nonetheless play a fundamental role in theorizing and are used to model information states and truth. In addition, many updates can be said to express propositions, however, certain epistemically modalized updates do not (Veltman, 1996, 231). Instead, they perform the eponymous test on an information state. Let an information state be a set of possible worlds compatible with some body of information. Updating with "It might be the case that p" tests the state for compatibility with p. If the state contains any worlds where p is true, then the initial state is returned (the test is passed). If there are no p worlds then $\varnothing$ is the output (the test is failed), (Veltman, 1996, 228).

We'll now take a look at dynamic semantics in a fairly general form. In particular, we consider the propositional fragment of the dynamic language proposed in Groenendijk and Stokhof (1991b). Several variants of this semantics exist, but this one possesses several crucial features that they have in common while also offering enough semantics machinery to address our puzzles. This is intended to be fairly introductory, and later chapter include greater formal and philosophical detail.

Let $\mathcal{L}_1$ be generated by the grammar:

$$\varphi ::= \mathrm{p} \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi$$

Dynamic semantics is intended to capture the manner in which sentences can update

[21] See Groenendijk and Stokhof (1991a) van der Does et al. (1997) Beaver (2001), Aloni (2001), Gillies (2004), von Fintel and Gillies (2007), Dever (2013), Yalcin (2012b), Yalcin (2015), Willer (2013), and Lennertz (2019) among others, for developments and applications of the test semantics. For influential dynamic accounts that do not directly concern epistemic modals, see Kamp (1981), Heim (1983), and Groenendijk and Stokhof (1991b). See Goldstein (2019) for an update semantics where not all instances of the epistemic modal operator perform a test.

a state of information based on their compositional structure. We thus characterize a general framework for the way in which sentences can update states of information. This can be captured by an *update system*.

(23) **Update System**: $\langle \mathcal{L}, C, \cdot[\cdot] \rangle$ is an update system if and only if $\mathcal{L}$ is a set of sentences, C is a set of information states, and $\cdot[\cdot]$ is a function which maps sentences of $\mathcal{L}$ to operations on C.

An update system models the manner in which sentences of a given language can change a state of information. Let W be the set of all functions $\mathcal{A} \mapsto \{0, 1\}$. Our update system will be $\langle \mathcal{L}_1, \mathcal{P}(W), \cdot[\cdot] \rangle$. An information state $s \in \mathcal{P}(W)$ is a set of possible worlds compatible with the information contained within the state. We characterize our semantics for $\mathcal{L}_1$ in terms of operations on $\mathcal{P}(W)$.

(24) **Semantic Clauses**[22]

1. $s[\mathrm{p}] = \{w \in s | w(\mathrm{p}) = 1\}$
2. $s[\varphi \wedge \psi] = s[\varphi][\psi]$
3. $s[\neg\varphi] = s - s[\varphi]$
4. $s[\Diamond\varphi] = \{w \in s : s[\varphi] \neq \varnothing\}$

Updating a state with an atomic formula takes the intersection of the worlds in the state, and the set of worlds where that atomic is true. Conjunction is consecutive update, and negation takes the difference between the initial state, and what would result by updating the initial state with the non-negated formula. Epistemic possibility (epistemic "might") performs a test, where the initial state is returned if the state is compatible with the prejacent of the modal. Otherwise, the test is "failed" and the absurd state $\varnothing$ is returned.

We can next define what it is for a state to support some information.

[22]The present formulation only considers negation, conjunction, and epistemic possibility operators. Disjunction and quantification are suppressed, as the problem can be formulated in their absence.

(25) **Support**: $s \vDash \varphi \Leftrightarrow s[\varphi] = s$

A state supports some information if and only if updating the state with that information does not change the state. In dynamic frameworks, belief is often defined in terms of support.[23] Next, we define consequence and consistency.

(26) **Dynamic Consequence**: $\psi_1, ..., \psi_n \Vdash \varphi \Leftrightarrow \forall s, s[\psi_1]...[\psi_n] \vDash \varphi$

(27) **Dynamic Consistency**: $\psi_1, ..., \psi_n$ is consistent $\Leftrightarrow \exists s, s[\psi_1]...[\psi_n] \neq \varnothing$

(28) **Dynamic Inconsistency**: $\psi_1, ..., \psi_n$ is inconsistent $\Leftrightarrow \psi_1, ..., \psi_n$ is not consistent

A sequence of formula dynamically entails some formula φ if and only if every state updated with that sequence will support φ. Similarly, a sequence of updates is dynamically consistent if and only if there is a state that can be updated with that sequence without resulting in the absurd state. Note that this notion of entailment does not concern truth, and instead captures support preservation. Also note that the above notions are order sensitive. For instance, $\langle \Diamond p, \neg p \rangle$ is dynamically consistent while $\langle \neg p, \Diamond p \rangle$ is not.

While we've elected to define dynamic entailment and consistency as above, non-equivalent alternatives are both available and useful. A variant of inconsistency that is interdefinable with a different notion of entailment is also often appealed to.[24]

(29) **Coherence**: $\psi_1, ..., \psi_n$ is coherent $\Leftrightarrow \exists s \neq \varnothing$, s.t. $s \vDash \psi_1, ..., s \vDash \psi_n$

Unlike consistency, coherence demands that a single state support each update individually, rather than in sequence. Unlike consistency, coherence is not order sensitive, e.g. neither $\langle \Diamond p, \neg p \rangle$ nor $\langle \neg p, \Diamond p \rangle$ are coherent. Any update that is not coherent (incoherent) is not supported by any single state.

[23] See Chapters 2 and 4 for discussion.

[24] See Chapter 4 for discussion of three distinct notions of dynamic entailment.

With this foundation, we can now return the puzzles of modal disagreement and epistemic contradiction, and see how dynamic semantics fares.

1.6 Dynamic Solutions

The first feature of dynamic semantics to observe is that despite the fact that the sequence $\langle \neg\varphi, \Diamond\varphi \rangle$ is dynamically inconsistent, it is not the case that epistemic "might" is factive: $\Diamond\varphi \not\models \varphi$. Thus, by embracing a non-classical entailment relation, the dynamic semanticist avoids the factivity of "might" while embracing the inconsistency of the conjuncts of epistemic contradictions.

We now return our attention to modal disagreements. Consider the conversation between Holmes and Watson. Watson believes $\Diamond m$ while Holmes believes $\neg m$. Observe that In virtue of believing $\neg m$ Holmes is incapable (without first revising his information state) of also believing $\Diamond m$. This is because any state that supports $\neg m$ also supports $\neg\Diamond m$ and cannot be updated with $\Diamond m$ without crashing. This is to say that the content of Watson's belief/utterance is dynamically inconsistent with the content of Holmes' belief/assertion.[25] The dynamic theorist may thus explain modal disagreements in terms of dynamic inconsistency. Dynamic semantics does not stipulate any particular modal base for a given instance of an epistemic modal. Rather the test that the modal initiates is applied to whatever information state it is updating. This allows for the exact same explanation to be applied to eavesdropper and temporal instances of modal disagreement as well. Importantly, this explanation does not merely apply to modal disagreements, but paradigm instances of disagreement too. For any atomic sentence p, p will be inconsistent with $\neg p$, and thus, paradigm instances of disagreement can be explained by exactly the same mechanism.[26] Thus,

[25] Technically speaking, *inconsistent with* is a two-place relation. While I take my meaning to be intuitive, we can more formally say that φ is inconsistent with $\psi \Leftrightarrow \langle \psi, \varphi \rangle$ is dynamically inconsistent. See Lennertz (2019) for a dynamic view which uses dynamic inconsistency to characterize instances of *asymmetric disagreement*. More on this below, and in Chapter 2.

[26] There is a substantial wrinkle here, as not every formula φ is inconsistent with its negation $\neg\varphi$.

the dynamic semanticist not only predicts and explains modal disagreements, but is able to do so in precisely the same way as she predicts and explains paradigmatic disagreements: in terms of dynamic inconsistency.

Dynamic inconsistency is similarly helpful in predicting the infelicity of epistemic contradictions. Consider, first, epistemic contradictions like (6) and (9) where the right conjunct is modalized. These have the following forms, respectively:

(30) $r \wedge \Diamond \neg r$

(31) $\neg r \wedge \Diamond r$

The infelicity of right modalized epistemic contradictions is easily predicted by the fact that $r \wedge \Diamond \neg r$ and $\neg r \wedge \Diamond r$ are contradictions according to dynamic semantics. Any state updated with either formula will result in the absurd state $\varnothing$. Hence, each are predicted to be infelicitous. What's more is that since this infelicity is explained semantically, it can also be used to explain the infelicity of right modalized epistemic contradictions embedded under attitude verbs, conditionals and suppositions. Importantly, this strategy does not require a particular commitment to the semantics of attitude verbs, conditionals or suppositions.[27]

Left modalized epistemic contradictions are little more nuanced. An earlier version of dynamic semantics from Veltman (1996) observes the discrepancy in felicity between the following two discourses:

(32) ?? Adam might be at the door... No, he isn't.

(33) # Adam isn't at the door... he might be.

Indeed, the law of of non-contradiction is invalid in dynamic semantics. I discuss this at length in Chapter 3.

[27]When contradictions are embedded under attitude verbs, suppositions, and in the antecedents of conditionals, utterances thereof tend to be judged as infelicitous, even if the meanings of the sentences are not absurd. We may thus appeal to such a principle in order to explain the infelicity of attitude verbs, suppositions, and conditionals with these features, without discussing all of the details of their semantics. Dynamic accounts of belief are discussed in Chapters 2 and 4. I largely avoid discussion of dynamic theories of conditionals in this dissertation, as the problems I address do not fundamentally involve them. However, see Gillies (2004) for a dynamic account of conditionals.

Veltman observes that discourses like (32) can be uttered with marginal felicity, while the opposite ordering in (33) cannot. Dynamic semantics was intended, in part, to predict this discrepancy. In dynamic semantics, the sequence of updates in (32) is consistent, while the other ordering in (33) is not. It is easy to see the parallels between discourses like those above, and epistemic contradictions, especially when conjunction is seen as consecutive update. The data collected in Knobe and Yalcin (2014) can be seen as a precisification of Veltman's intuition. Here it was observed that epistemic contradictions with left modalized conjuncts where still judged to be infelicitous, but only marginally so in comparison to right modalized epistemic contradiction.

The puzzle for the dynamic semanticist is to explain the marginal infelicity of left modalized epistemic contradictions. This can be done by appeal to an alternative definition of consistency called coherence. Sentences like $\Diamond a \wedge \neg a$, while dynamically consistent, are not coherent. Appeal to their incoherence can be used to explain their marginal infelicity. It is important to appreciate that this explanation is not *ad hoc*. Updates with left modalized epistemic contradictions like $\Diamond a \wedge \neg a$ can often occur, and may be infelicitous, however, they are not infelicitous in a way that is catastrophic. Since they are dynamically consistent, the conversation in which they occur can sustain update with them without yielding a catastrophic failure of the context set, much the same as (32). Despite this, they are incoherent in that no single state can support all of the information contained in the update. Similarly, we can plausibly explain the infelicity of embedded left modalized epistemic contradictions by appealing to their incoherence, in precisely the same way that we predicted the infelicity of embeddings that were inconsistent.[28] We can also see that while both right and left modalized epistemic contradictions are, in some sense, deviant, right modalized epistemic contradictions are worse, insofar as update with them is always

[28]This strategy for solving modal disagreements and epistemic contradictions in a dynamic framework can be found in Willer (2013).

catastrophic, since they are contradictions.

Thus, the dynamic semanticist is able to predict and explain modal disagreements, while similarly predicting the infelicity of epistemic contradictions. What may be less clear is how this is done in the unified fashion that I suggest. According to the story above, modal disagreements are explained in terms of dynamic inconsistency, but some epistemic contradictions are predicted to be infelicitous due to their incoherence. The dynamic theorist has two options for pursuing a unified account, should they choose to.

1.6.1 Incoherence First Approach

The first option available to the dynamic semanticist is to take incoherence as the predictor of disagreements and semantic infelicity. This is easy to do since any sequence of sentences that is dynamically inconsistent will yield a set of sentences that is incoherent. Previously, we explained disagreements in terms of dynamic inconsistency, but we may instead choose to associate disagreement with incoherence. This can be captured by the following principle.

(34) **Disagreement as Incoherence**: agents a and b disagree only if a believes/utters some content φ and b believes/utters some content ψ and $\{\varphi, \psi\}$ is incoherent.[29]

Characterizing disagreement in terms of incoherence strikes as intuitive and seems to be motivated by intuitions similar to those that motivate simultaneous belief.[30]

[29]So far, I have been loose with the distinction between two notions of disagreement. The first treats disagreement as a relationship between two agents that holds in virtues of their mental states. Call this the state based account. The second treats disagreement as an action that an agent performs, by rejecting the speech act of the other. The puzzles, as I have currently characterized the, do not rely on either characterization, and so I'll be loose with the distinction here. In later chapters, however, this will become more important. See Chapter 2 for discussion.

[30]Indeed, one need not even embrace dynamic semantics in order defend disagreement as incoherence. Yalcin (2007) and Yalcin (2011) adopt disagreement as coherence in a static framework.

What's more, is that this explanation, similar to the explanation that involved dynamic inconsistency, is able to predict paradigmatic instances of disagreement as well.

We can then apply a similar principle for the infelicity of contradictions, epistemic and otherwise.

(35) **Incoherent only if Infelicitous**: If a discourse or sentence is incoherent, then utterance thereof is infelicitous.

The standard approach is to treat inconsistency as indicative of infelicity, i.e. if an agent utters a sentence or series of sentences that are inconsistent, that utterance should be infelicitous. The dynamic strategy considered previously was to replace classical inconsistency with dynamic inconsistency and apply the very same approach. However, we might instead insist that incoherence is a predictor of infelicity. This isn't much of a leap since everything that is dynamically inconsistent is already incoherent. Thus, both left and right modalized epistemic contradictions, as well as paradigmatic contradictions, are incoherent and predicted to be infelicitous by the principle above. This strategy has numerous advantages. In the first place it, it allows us to predict the infelicity of contradictions, epistemic and, paradigmatic, with a single principle. We can additionally suggest that utterances with dynamically inconsistent contents will have increased infelicity, explaining the discrepancy between left and right modalized epistemic contradictions. Secondly, it dovetails nicely with disagreement as incoherence to yield a unified explanation of modal disagreements and epistemic contradiction. According to this explanation, all disagreements are explained by incoherence, and the infelicity of both contradictions and infelicitous non-contradictions can similarly be explained in terms of incoherence. Thus, appeal to incoherence not only provides a single mechanism for explaining modal disagreements and epistemic contradictions, but it also explains how each are related to paradigmatic disagreements and contradictions. According to this strategy, the unifying thread is incoherence.

1.6.2 Asymmetric Approach

Dynamic semantics is not limited to the above explanation. Indeed, a secondary explanation becomes attractive upon a modified construal of modal disagreements. I have suggested that modal disagreements and epistemic contradictions are related puzzles that demanded a unified explanation. There is, however, something of an discrepancy in the puzzles, specifically with respect to the graded infelicity of right vs left modalized epistemic contradictions. At least as I have described the puzzle thus far, there is no parallel to this phenomenon in modal disagreements, and one may argue that this discrepancy is enough to justified unique solutions to each puzzle.

However, upon further inquiry, parallels can be found. The puzzle of modal disagreement as I have characterized it, and how it is typically characterized, asks whether Holmes disagrees with Watson or whether there is a disagreement between Holmes and Watson. Less often asked is whether Watson disagrees with Holmes. This is typically because classical inconsistency will always predict a symmetric disagreement relation, e.g., Holmes disagrees with Watson if and only if Watson disagrees with Holmes. However, there is reason to suspect that Holmes disagrees with Watson, but Watson does not disagree with Holmes. Observe that normative judgment holds in Holmes' case, as he is licensed to assert that the content of Watson's belief/utterance is wrong. However, in Watson's case, normative judgment fails. We may stipulate that Watson is undecided as to who the murderer is. This is to say, he does not know whether m or $\neg m$. He is thus not licensed to judge the content of Holmes' assertion as wrong or otherwise incorrect. He can normatively evaluate whether or not Holmes is justified in his utterance, but he cannot normatively evaluate the content of that utterance, which is the target of normative assessment.[31] Should we think that nor-

[31] In Chapter 2, I make this distinction in greater detail by introducing the notions of: *formative correctness* (whether or not one is justified in forming a belief), and *content correctness* (whether or not the content of the belief is, itself, correct. This distinction is terminologically different, but fundamentally the same as a distinction discussed in Lennertz (2019).

mative assessment is not merely indicative of *a* disagreement, but indicative of *who* disagrees with *whom*, then we may be inclined to think that the disagreement between Holmes and Watson is asymmetric. This idea is proposed in Lennertz (2019).

Should we choose to view the puzzle this way, the parallels with the graded infelicity of epistemic contradictions return. Upon this construal, modal disagreements and epistemic contradictions are each asymmetric phenomena, and the similarity of the puzzles is preserved. Dynamic semantics is uniquely situated to explain this and can do it by adopting the following thesis.

(36) **Disagreement as Dynamic Inconsistency**: agent a disagrees with agent b only if a believes/utters some content φ and a believes/utters some content ψ and φ is inconsistent with ψ.

On this account, dynamic inconsistency models who disagrees with whom. We see that Watson's claim of $\Diamond m$ is inconsistent with Holmes' claim of $\neg m$, so Holmes disagrees with Watson. However, Holmes' claim is not inconsistent with Watson's, and so Watson does not disagree with Holmes. According to this proposal, modal disagreements differ from paradigmatic disagreements insofar as they are asymmetric.[32] This asymmetry mirrors the asymmetry in epistemic contradictions, as left modalized epistemic contradictions are contradictions proper, while right modalized ones are not. We can explain their graded infelicity just as we did at the beginning of the section.. Viewed this way, both puzzles are asymmetric in nature, and thus, symmetry between the puzzles is preserved, and the same explanation involving both dynamic inconsistency and incoherence applies properly to both.

Thus, the dynamic semanticist has at least two ways to provide a unified solution to both puzzles. Not only do they have the freedom to construe disagreement

[32]The presence of asymmetric disagreement is unremarkable if we take disagreement to be an action; one can easily reject the claim of another without them rejecting yours. However, the thesis is highly nontrivial if we view disagreement as a state.

in ways that static theories do not, but unified solutions fall out fairly naturally.[33] The competing theories considered, in the best case, will have to contort themselves considerably to achieve the same success.

1.7 Trouble

I have argued that dynamic semantics is able to provide a unified solution to the puzzles of modal disagreement and epistemic contradiction, while relativism and bounded semantics cannot. I take this to count as a legitimate point in favor of dynamic semantics. However, the picture I've painted thus far, like those by many dynamic semanticists before me, is strategically rosy. The tone henceforth substantially shifts. Dynamic semantics is able to achieve its ends by rejecting both classical consequence and canonical theories of content. These deviations from the canon lose access to numerous plausible and attractive strategies that allow canonical accounts to interact with various theses concerning epistemology, and normativity. This comes with several substantial explanatory burdens, many of which remain irresponsibly unanswered and in many cases unrecognized. While dynamic semantics can successfully solve a narrow and curated list of puzzles, deeper questions loom that that truth-conditional theories do not face. This dissertation considers three such problems with a chapter dedicated to each.

The first problem, addressed in Chapter 2, arises from the fact that dynamic semantics treats contents as CCPs rather than truth-conditions. In many cases, truth-conditions can be recovered from CCPs, and such contents can be, at least indirectly, treated as true and false. Other contents, however, cannot be associated

[33]Other theories can also provide unified solutions, and I take my arguments support them in the same way they support dynamic semantics. Two fairly prominent accounts that do this are Yalcin's domain semantics, defended in Yalcin (2007) and Yalcin (2011), as well as Moss' probabilistic semantics, defended in Moss (2015) Moss (2018). Not coincidentally, Yalcin's proposal, while static, takes a great deal of inspiration from dynamic theories. Moss' proposal is dynamic, but takes contents to be sets of probability spaces. Without too much squinting, one can see it as a development of the dynamic semantics considered here.

with any proposition, and thus, cannot even indirectly be associated with notions of truth and falsity. This yields difficult problems when it comes to our normative judgments about these contents. Indeed, the few extant proposals regarding the normativity of dynamic content fail under minimal pressure. This is significant, as the disagreement challenge, which dynamic semantics aims to solve, is motivated by the same normative judgments that dynamic semanticists have failed to explain. In order to avail themselves of the above solutions, dynamic semanticists must provide an account of correctness and incorrectness that is appropriate for non-truth conditional content, while also properly motivates the puzzle they aim to solve. I strive to fill this lacuna by providing an account of the normativity of dynamic content that satisfies these criteria.

Chapter 3 considers a problem concerning the non-classical behavior of dynamic entailment. While, dynamic semantics has no shortage of non-classical behavior, this behavior is only selectively explained. In particular, dynamic semantics rejects the law of Non-Contradiction, with no explanation of which I am aware. Critics, most notably Mandelkern (2020) consider this to be a failure of dynamic semantics. I defend and explain failures of Non-Contradiction by comparing dynamic semantics and classical, truth-conditional semantics in terms of their idealizing assumptions. I demonstrate that dynamic semantics rejects context fixity, an idealizing assumption which truth-conditional semantics adopts. I then argue that any semantics which rejects context fixity should, by the classical semanticists own lights, violate non-contradiction under certain circumstances. I then demonstrate that dynamic semantics violates Non-Contradiction in all and only those circumstances. I then consider further indirect evidence in support of this result. I close by suggesting that discussion of idealizing assumptions, common in the sciences, is similarly crucial to fruitful discussion in natural language semantics.

Chapter 4 introduces a novel epistemic puzzle that does not occur in static frame-

works. The puzzle involves contents which are decisive yet non-committal. Content is decisive with respect to an issue if and only if an agent who believes that content must be decided on the issue. Content that is non-committal is content that does not commit the believing agent to a particular decision, even if they are decided. Dynamic semantics allows for the unique case where contents can be both decisive yet non-committal, i.e. if an agent believes such content, they must be decided, but the content itself does not commit the believing agent to a particular choice on the issue. This results in an epistemic puzzle whereby undecided agents can update with decisive, non-committal content, and after update, are decided. To solve the puzzle, I appeal to a long forgotten notion of entailment, introduced in Veltman (1996), and largely ignored thereafter, called minimal entailment. After demonstrating how and why minimal entailment allows the dynamic semanticist to solve the puzzle, I recommend that dynamic semanticists more fully embrace the extant practice of utilizing multiple entailment relations within the same semantic framework.

With respect to each of the above, I take recognition of the problem to be far more substantial than my proposed solutions. If dynamic semantics is to be taken seriously, then these and other philosophical issues must be addressed.

Chapter 2

Dynamic 'Might' and Correct Belief

Synopsis: Veltman's *test semantics* and developments thereof reject the canon about semantic contents and attitude ascriptions in favor of dynamic alternatives.[1] According to these theories the semantic content of a sentence uttered in context is not a proposition, but a context change potential (CCP). Similarly, beliefs are not taken to be relations between agents and propositions, but agents and CCPs. These deviations from the canon come at the cost of an elegant explanation about the correctness of belief. Standardly, it is taken that the content of a belief is correct just in case the content of that belief is true. The proponent of the test semantics cannot appeal to this explanation since they hold that certain contents, namely epistemically modalized contents, do not express propositions, and are neither true nor false. Willer (2013) concerns how the test semantics can be martialled to solve *inter alia* puzzles of modal disagreement. Crucial to Willer's account is the proposal of a correctness condition which I call *evidential correctness*. According to evidential correctness, the content of

[1]The test semantics is introduced in Veltman (1996). See Groenendijk and Stokhof (1991a), Groenendijk and Stokhof (1991b), and van der Does et al. (1997) among many, many others for further developments.

a belief is correct just in case someone with the believing agent's evidence would not be committed to factual error. The present paper argues, contra Willer, that evidential correctness does not yield the proper correctness judgments for interlocutors in several common instances of modal disagreement. Since these correctness judgments are what motivate the claim that modal disagreements are genuine disagreements, I take the objection to be significant. I subsequently consider two potential alternatives: *weak correctness* and *strong correctness* but conclude that each suffers from substantial problems. I then provide my own novel correctness conditions for belief contents within a dynamic framework. I argue that there are, in fact, two conditions for correctness of content that ought to be considered when operating within the test semantics. The first, *locative correctness*, applies only to contents which locate the believing agent in some subset of the space of possibility by entailing some contingent proposition. The second, *informational correctness* applies to contents that do not not locate the believing agent. Such content includes, most notably, epistemically modalized content. After motivating this distinction, I demonstrate how it yields the requisite correctness judgments on the part of disagreeing agents, and avoids the problems of the previously considered views.

Introduction

Canonical accounts of semantic content hold that the semantic value of a sentence is a proposition. Canonical accounts of belief hold that belief is a relation between an agent and a proposition. Thus, the canon often helps itself to an elegant explanation about the correctness of belief: a belief is correct if and only if the content of that belief is true, (Wedgewood, 2002, 267). Hence the slogan "belief aims at truth."[2] A popular, dynamic, story about epistemic modals, often called *update*

[2]See Williams (1973) for the earliest instance of this slogan.

semantics or *test semantics*, rejects the canon about content.[3] Introduced in Veltman (1996), and developed extensively thereafter, the test semantics has it that contents are not static propositions which are true or false. Instead, they are dynamic updates–functions from contexts to contexts–that provide instructions on how to update a body of information. Thus, meanings are not propositions but context change potentials. Propositions nonetheless play a fundamental role in theorizing and are used to model information states and truth. In addition, many updates can be said to express propositions, however, certain epistemically modalized updates do not (Veltman, 1996, 231). Instead, they perform the eponymous test on an information state. Let an information state be a set of possible worlds compatible with some body of information. Updating with "It might be the case that p" tests the state for compatibility with p. If the state contains any worlds where p is true, then the initial state is returned (the test is passed). If there are no p worlds then $\varnothing$ is the output (the test is failed), (Veltman, 1996, 228). Importantly, agents can have epistemically modalized beliefs, and if the contents of these beliefs are not propositions, then belief cannot be characterized as a relation between an agent and a proposition. Thus, proponents of the test semantics also reject the canon about belief. Instead, dynamic accounts often adopt an alternative defended in Heim (1992), according to which, belief is defined as the fixed point of an update. An agent believes some content just in case their information state remains unchanged when updated with that content. The relevant advantage of this view of belief is that the associated content need not be a proposition, and thus, agents can hold non-propositional beliefs.

So far, so good, but rejecting the canons comes at an indirect cost. Notably, the proponent of the test semantics cannot appeal to the elegant explanation about

[3]See Groenendijk and Stokhof (1991a) van der Does et al. (1997) Beaver (2001), Aloni (2001), Gillies (2004), von Fintel and Gillies (2007), Dever (2013), Yalcin (2012b), Yalcin (2015), Willer (2013), and Lennertz (2019) among others, for developments and applications of the test semantics. For influential dynamic accounts that do not directly concern epistemic modals, see Kamp (1981), Heim (1983), and Groenendijk and Stokhof (1991b). See Goldstein (2019) for an update semantics where not all instances of the epistemic modal operator perform a test.

correct belief mentioned above. Since it is held that not all beliefs are true or false, their correctness cannot be explained exclusively in terms of truth and falsity. Instead, dynamic theorists appeal to correctness conditions that are related to truth only indirectly. Willer (2013) provides what is, to my knowledge, the only dedicated account of correctness of belief for the test semantics. Call this *evidential correctness*. According to evidential correctness, a belief is correct if and only if it would not commit someone with the believing agent's evidence to factual error, (Willer, 2013, 64). Crucially, this proposal does not require that the belief itself be true or false.

Willer (2013) develops the test semantics in order to solve puzzles of *modal disagreement*. Modal disagreements paradigmatically involve an agent a who believes that some proposition p is not the case and some agent b who believes that p might be the case. It is claimed that a disagrees with b, and the fact that the test semantics treats these contents as inconsistent is touted as a motivation to adopt the view. The puzzle is often marshalled against the contextualist semantics for epistemic modals because contextualism allegedly fails to explain the disagreement. It is further argued that modal disagreements are genuine disagreements *because* a is disposed to judge the content of b's belief as wrong, incorrect, or mistaken.[4] This is supported by the observation that a can respond with something like, "No, you're wrong," or "No, that's not right," in response to b's assertion.[5] This is to say that a is disposed to judge b's belief as incorrect, and that this judgement supports the claim that a disagrees with b.

The present paper argues that evidential correctness fails to yield the proper correctness judgments in a wide variety of cases. In short, if a does not share evidence with b, and a makes correctness judgments based on evidential correctness, then there will be cases where a will not conclude that the content of b's belief is incorrect since b's

[4]Henceforth, I shall use normative terms like "wrong", "incorrect", etc. interchangeably.

[5]Appeal to correctness judgments that motivate disagreement can be found in MacFarlane (2011), Willer (2013), and Lennertz (2019).

belief that it might be that p would not commit anyone with b's evidence to factual error. a, nonetheless, disagrees with b. Unfortunately, many paradigmatic modal disagreement scenarios concern cases precisely like this, and evidential correctness fails to provide the motivating normative judgment on the part of a. This is a problem, since the puzzle of modal disagreement is, itself, motivated by the claim that a judges that the content of b's belief is incorrect. Since these correctness judgments motivate the problem of modal disagreement, proponents of the test semantics must provide a tenable alternative that yields the appropriate judgment, i.e. a must consider the content of b's belief to be incorrect in all cases of modal disagreement. Otherwise, the disagreement challenge loses its teeth.

I next consider two possible alternatives for correctness of belief: *weak correctness* and *strong correctness*. I conclude, however, that despite their advantages, both are implausible. I then propose a distinction between two kinds of correctness that govern the different types of content within the test semantics. Call these two kinds of correctness *locative correctness* and *informational correctness*. In brief, content that accurately individuates the way the world is from the way that it is not, is locatively correct. This is to say that such content locates the believer within a contingent set of worlds which contains the actual world. Similarly, content that misrepresents the way the world is is locatively incorrect. However, beliefs that do not locate the believer in one world or another, e.g. "might" beliefs, neither represent nor misrepresent the world. Thus, they are neither correct nor incorrect, locatively speaking. Such contents, I argue, are instead beholden to the norm of *informational correctness*. Some content is informationally correct just in case that content is consistent with the information that represents the facts. Thus, instead of judging the correctness of epistemically modalized contents with respect to their capacity to represent (which they cannot do), they are judged by their capacity to guide inquiry towards the truth (which they can do). The present paper develops these notions in a way that

is consistent with the correctness judgments that motivate the test semantics. In addition, I briefly explain why this doxastic framework is not only plausible, but preferable, in that it distinguishes between the manner in which ignorance fails to be correct, and the manner in which misrepresentation fails to be correct. I then show that the framework makes plausible predictions with respect to how we ought to believe under uncertainty, and how we utilize doxastic norms to guide our doxastic actions.

§1 introduces the puzzle of modal disagreement and briefly rehearses the test semantics. §2 develops evidential correctness, weak correctness, and strong correctness. Problems are posed for each, and none are considered satisfactory. §3 considers some possible lines of response to the challenges posed in §2 but concludes that none are successful. §4 presents and explains locative and informational correctness then demonstrates how adopting both avoids the problems of weak, strong, and evidential correctness. §5 closes with some brief remarks.

2.1 Preliminaries

2.1.1 Modal Disagreement

Consider the following well-rehearsed scenario: Holmes believes that Moriarty is not the murderer ($\neg m$). Watson is undecided, and believes both that Moriarty might be the murderer (Might m) and that he might not be (Might $\neg m$).[6] Numerous criticisms of the canonical *contextualist* semantics for epistemic modals rest on the claim that in scenarios like this, Holmes *disagrees* with Watson.[7] Call such cases *modal disagreements*.[8] The idea is that disagreement manifests in cases where the disagreeing agents

[6]Here, and henceforth, the modal auxiliary "might" is to be interpreted epistemically.

[7]See Kratzer (1977) and Kratzer (1981) for characterizations of contextualism.

[8]Several variations of this scenario exist. This version is structurally identical to the one from Lennertz (2019) which characterizes disagreement as a relation between two believing agents that holds in virtue of their belief states. Upon this construal, disagreement is treated as a state, rather

hold beliefs that are inconsistent with each other's. According to contextualism, $\neg m$ and Might m are consistent, and the disagreement goes unexplained.[9] Crucially, the pressure to consider such cases as genuine disagreements is often motivated by the observation that Holmes believes, and is licensed to assert, something like "No, you're wrong," directed at Watson's belief (or assertion) that Moriarty might be the murderer.[10] Similar scenarios involving diachronic intrapersonal disagreement have also been considered, where an agent who at one point believed that p might be the case, and later learns that p is not the case, can remark "I used to believe that p might be the case, but I was wrong."[11] In each case, some agent holds a belief, and in virtue of holding that belief, they make a normative judgment that some other belief (and the agent who holds it) is wrong. This judgment is treated as evidence that the first agent disagrees with the second. Contextualism's alleged failure to properly predict these instances of disagreement is often used to motivate various non-canonical alternatives, including the test semantics.

2.1.2 Test Semantics

We now consider the test semantics in a simple and fairly general form. More sophisticated variants, most notably the semantics in Willer (2013), will be discussed in §3.[12]

Let $\mathcal{L}_1$ be generated by the grammar:

$$\varphi ::= \mathrm{p} \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \Diamond\varphi$$

than an action, and disagreements need not manifest directly in conversation. For variants of the puzzle, see Egan et al. (2005) Willer (2013), MacFarlane (2011) and MacFarlane (2014). See Huvenes (2017) for more on disagreement as a state.

[9] See von Fintel and Gillies (2011) for a defense of contextualism. See Lennertz (2020) and Yalcin (2015) for responses. Also see Plunkett and Sundell (2013).

[10] See MacFarlane (2011) and Lennertz (2019).

[11] See MacFarlane (2011), Lennertz (2019), and Willer (2013).

[12] The following formulation is equivalent to the one from Lennertz (2019) which is inspired by Veltman (1996).

Definition 1: Update System

$\langle \mathcal{L}, C, \cdot[\cdot] \rangle$ is an update system if and only if $\mathcal{L}$ is a set of sentences, C is a set of information states, and $\cdot[\cdot]$ is a function which maps sentences of $\mathcal{L}$ to operations on C.

Let W be the set of all functions $\mathcal{A} \mapsto \{0, 1\}$. Our update system will be $\langle \mathcal{L}_1, \mathcal{P}(W), \cdot[\cdot] \rangle$. If s is a state (a set of worlds in $\mathcal{P}(W)$) and φ is formula of $\mathcal{L}_1$, then $s[\varphi]$ is the result of updating s with a formula φ. We now define the relevant notions of support and entailment, in addition to the semantic clauses for the connectives.

Definition 2: Support

$s \vDash \varphi \Leftrightarrow s[\varphi] = s$

A state supports a formula just in case updating the state with that formula results in the original state. This is to say that a state supports a formula just in case the information in that formula is already contained within the state.

Definition 3: Semantic Clauses

1. $s[\mathrm{p}] = \{w \in s | w(\mathrm{p}) = 1\}$
2. $s[\varphi \wedge \psi] = s[\varphi][\psi]$
3. $s[\varphi \vee \psi] = s[\varphi] \cup s[\psi]$
4. $s[\neg\varphi] = s - s[\varphi]$
5. $s[\Diamond\varphi] = \{w \in s : s[\varphi] \neq \varnothing\}$

Update with an atomic sentence removes any worlds in the initial state where that sentence is not true. Conjunction is consecutive update, while disjunction is the union of the individual updates of the disjuncts. Negation removes the result of updating with the formula negated. Might φ runs a test on the information state, and returns

the input state if update with φ doesn't crash. Otherwise, it crashes, (Veltman, 1996, 228). We can also define a Must operator as the dual of Might.[13]

Definition 4: Dynamic Consequence

$\psi_1, ..., \psi_n \Vdash_D \varphi \Leftrightarrow \forall s, s[\psi_1]...[\psi_n] \vDash \varphi$

A sequence of formula dynamically entails another formula just in case every state is such that sequential update with those formula will support the other formula. This notion of consequence is often called *update-to-test consequence*.[14] This definition of consequence is interdefinable with the following notion of consistency.

Definition 5: Dynamic Consistency

$\psi_1, ..., \psi_n$ is consistent$_D \Leftrightarrow \exists s, s[\psi_1]...[\psi_n] \neq \varnothing$

Definition 6: Dynamic Inconsistency

$\psi_1, ..., \psi_n$ is inconsistent$_D \Leftrightarrow \psi_1, ..., \psi_n$ is not consistent$_D$

A sequence of formula is dynamically consistent just in case there is some state that can sustain the sequence of updates without resulting in the absurd state. Otherwise, it is dynamically inconsistent. Relevantly, dynamic consequence and consistency are order sensitive. For instance, the sequence Might p, $\neg p$ is consistent$_D$ while the sequence $\neg p$, Might p is inconsistent$_D$.

Even though semantic contents in a dynamic framework are not characterized in terms of propositions, propositions still do the representational heavy lifting, and many updates express propositions. The characteristic feature of updates like Might p is that they do not express propositions, and thus, we cannot even indirectly speak of their truth and falsity or their associated facts. Moreover, it is well known that the

[13]Must $\varphi =_{\text{def}} \neg$Might $\neg\varphi$. On such an interpretation, Must is strong. While there are reasons to reject the strength of English "must" (see Karttunen (1972)), many working in the test semantics take "must" to be strong, so I shall follow suit. See von Fintel and Gillies (2007), Willer (2013), and Lennertz (2019) for such implementations. See von Fintel and Gillies (2007) and von Fintel and Gillies (2010) for arguments in favor of this position.

[14]Veltman (1996) calls this notion of entailment $\Vdash_2$.

non-modal fragment of the semantics above behaves classically, but introduction of the Might operator introduces non-classical behavior. It will thus be useful to distinguish updates which express propositions from those that do not. Some content expresses a proposition when it has two properties: *distributivity* and *eliminativity*, (Groenendijk and Stokhof, 1991a, 57). All contents in the test semantics are eliminative, so some content is distributive if and only if it expresses a proposition.

Definition 7: Distributivity

φ is distributive $\Leftrightarrow$ for all $s, s[\varphi] = \bigcup\{\{w\}[\varphi] | w \in s\}$

An update is distributive just in case updating a state with the formula will yield the same result as the union of the individual updates of the singleton set of each world in that state, (van Benthem, 1989, 364). All contents that do not include the Might operator are distributive, and hence, express propositions. Paradigmatically, contents of the form Might p are not distributive, and do not express propositions.[15] Subsequent discussion will appeal to content that is propositional, which is to say, content that expresses update with a proposition.[16] It will sometimes be useful to refer directly to a proposition expressed by a distributive update, rather than the update itself. We capture this as follows:

Definition 8: Proposition

For any φ that is distributive, $[\![\varphi]\!] =_{\text{def}} \{w \in W | \{w\}[\varphi] \neq \varnothing\}$

Distributive contents can always be associated with a set of possible worlds or *proposition*. The $[\![\cdot]\!]$ allows us to refer to this proposition directly. Crucially, $[\![\cdot]\!]$ is only defined for distributive contents.

[15]Notably, not all formulas of the form Might φ fail to be distributive. For instance, when the Might operator takes wide scope over a contradiction, e.g. Might$(p \wedge \neg p)$, the formula will be distributive. As Mandelkern (2020) observes, however, not every formula of the form $\varphi \wedge \neg\varphi$ is a contradiction within the test semantics.

[16]I will occasionally say of some propositional content that that content is true (or false). This is bit of shorthand to say that the content expresses a proposition that is true (or false).

Lastly, we need to define what it is for an agent to hold a belief. Here I provide Heim's commonly accepted account of belief in dynamic frameworks.[17] In traditional static frameworks, such contents will be propositions, but this won't be so for dynamic theories. Instead, to have a belief is to have an information state that supports the update associated with the content of that belief.

Definition 9: Belief

An agent a believes some content $\varphi \Leftrightarrow I_a[\varphi] = I_a$

An agent holds a belief just in case update with the contents of that belief will not change the agent's information state. This is to say that the information contained in the update is already captured by the state, and thus, subsequent update has no effect. This can be equivalently defined in terms of support: an agent a believes some content φ if and only if $I_a \vDash \varphi$. Accordingly, an agent who's information state contains no p worlds will believe $\neg p$. Similarly, an agent who's information contains at least one p world believes Might p, and so forth. This account of belief is highly idealized insofar as agents who hold a belief will hold all the consequences of that belief. If an agent believes φ, and $\varphi \Vdash_D \psi$ then the agent also believes ψ. Similarly, to be undecided with respect to φ just is to believe Might φ and Might $\neg\varphi$.

The semantics above has been used to great effect to predict and explain various phenomena involving epistemic modals.[18] Here, however, we isolate our attention to modal disagreements. Holmes believes $\neg m$ because updating Holmes' information state, I_h, with $\neg m$ returns his initial state I_h ($I_h[\neg m] = I_h$). Watson believes that Might m and Might $\neg m$, since his information state I_w includes both m worlds and $\neg m$ worlds, and so update with either Might m or Might $\neg m$ (or both) will output I_w. However, should Holmes update his information state with Watson's belief that

[17] Instances of this can be found in Heim (1992), Yalcin (2011), Yalcin (2012b), Willer (2013) and Lennertz (2019) among others.

[18] Perhaps most notably, the test semantics predicts the infelicity of embedded epistemic contradictions. See von Fintel and Gillies (2007), Willer (2013), and Willer (2015) for details. See Yalcin (2007) for a thorough characterization of the problem of epistemic contradiction.

Might m, Holmes' state will crash. More formally, for any state s, if $s \vDash \neg m$ then $s[\text{Might } m] = \varnothing$. Holmes' belief that Moriarty is not the murderer has ruled out all m worlds. Thus, updating with Might m will check for m worlds, but is doomed to fail, since his belief that $\neg m$ ensures that there aren't any. Therefore, the content of Holmes' belief is inconsistent$_D$ with the content of Watson's belief, and Holmes thereby disagrees with Watson.

If disagreement is characterized in terms of inconsistency, the test semantics is able to predict that Holmes disagrees with Watson. Cases of diachronic intrapersonal disagreement can be explained in similar fashion. The issue, however, does not lie within the machinery of the test semantics. Rather, the alleged cases of disagreement that motivate the adoption of the test semantics are themselves motivated by the correctness judgments made by interlocutors like Holmes. This story is vindicated by the fact that Holmes judges the content of Watson's belief to be incorrect. The test semantics, by itself, does not tell us anything about what it means for the content of a belief to be correct. Thus, we need to supplement the formal machinery above with a theory of correctness that explains why Holmes makes this judgment.

2.2 Correctness Conditions

2.2.1 Formative Correctness vs. Content Correctness

Within the Kratzerian contextualist framework, Watson's belief that Might m expresses a second order description of the state he is in; something like, "For all I know, Moriarty might be the murderer." Thus, the content concerns Watson. If Watson is undecided as to whether Moriarty is the murderer, this content is true, and thereby correct. Moreover, this content is consistent with what Holmes believes. Unless Holmes misunderstands the nature of Watson's beliefs, Holmes cannot plausibly disagree with respect to the content of Watson's belief, nor can he believe that

content to be incorrect. Since contextualism does not predict the content of Holmes' belief and the content of Watson's belief as jointly inconsistent, it is argued that contextualism fails to predict these cases of disagreement. Contextualists, however, may respond that purported disagreements like those between Holmes and Watson are not genuine disagreements, or at least, not the kind of disagreements that concern contents and inconsistency therebetween. To ameliorate these worries and motivate the objection, it is common to invoke the correctness judgments of interlocutors. For instance, Holmes can respond or believe that the content of Watson's belief that Moriarty might be the murderer is wrong. The plausible assumption here is that when we judge the content of some belief or assertion to be incorrect, this is evidence of disagreement. The strategy of appealing to normative judgments about content strives to isolate the content of the belief or assertion as the locus of incorrectness, and thereby disagreement. This is to say that those who take the disagreement challenge to be an objection to Kratzerian contextualism insist that disagreement concerns content directly and place a demand on competing views that these disagreements manifest themselves in the semantics. This strategy places two constraints on proposals that can successfully answer the challenge. The first, and most well-recognized in the literature, applies pressure to commit to the claim that Holmes disagrees with Watson. The second is often less appreciated and holds that since the object of the normative judgement is the content of Watson's belief/assertion, it constrains the lines of response available to views which properly answer the challenge. In particular, it limits pragmatic and (arguably) metalinguistic explanations of the disagreement. Such restrictions are plausible, especially when we consider that disagreements tend to be *about* some content, and the disagreement in the present case is about whether or not Moriarty is the murderer. Thus, the correctness/incorrectness judgments of interlocutors like Holmes are a primary motivating factor for taking modal disagreements to be genuine disagreements, as well as insisting that they be explained by the con-

tents of beliefs. Unlike the Kratzerian semantics, the test semantics does predict the content of Holmes' belief to be inconsistent with the content of Watson's. However, in order to utilize Holmes' correctness judgments as motivation, proponents of the test semantics will need to provide correctness conditions that explain why Holmes judges the *content* of Watson's belief to be incorrect.

Before considering proposals for such conditions, it is worth observing precisely which kind of correctness will suffice. Lennertz (2019) makes a useful distinction in this regard, and I'll help myself to it here. Lennertz distinguishes between being correct/incorrect in believing some content and being correct/incorrect to believe some content. The first concerns whether the content of the belief is correct, while the second is an evaluation of how the belief was formed, (Lennertz, 2019, 4789).[19] To see the difference, suppose that p is true. Then suppose that a believes that p, while b believes, falsely, that $\neg p$. Suppose further that a is doxastically reckless and has developed their true belief on radically insufficient evidence. Moreover, b is rational, has been doxastically responsible, and b's best evidence supports $\neg p$. Thus, we can say that b is correct to believe $\neg p$, yet b is nonetheless incorrect in believing $\neg p$. The opposite holds for a, who is right in believing p but is wrong to believe p. While the process by which a formed the belief that p is unsatisfactory, and they are thereby wrong to believe p, the content of p is nevertheless correct, and a is correct in believing p.

Thus, there are two ways to normatively judge an agent with respect to their doxastic state. One concerns the process by which the agent formed their belief. We might call this *process correctness* or *formative correctness*. An agent is formatively correct to believe some content iff their belief forming process satisfies the appropriate evidential and doxastic standards, whatever they may be. The other type of correct-

[19]Note that this distinction is not restricted to dynamic accounts and is equally applicable to static accounts. For instance, Wedgewood's thesis clearly concerns correctness in believing, and does not concern correctness to believe.

ness, and the one relevant for disagreement, is determined wholly by the content of the belief. An agent is correct in believing some content iff that content is correct. The former notion of correctness concerns the inferential actions of the agent, and is sensitive to their evidence. The latter is not, and is entirely determined by the content of the belief. While both notions of correctness can also be applied to agents, formative correctness is sensitive to agential features over and above the content of the belief. On the other hand, whether an agent is correct in believing some content is determined solely by the content of the belief in question. As Lennertz argues, it is the second kind of correctness that is relevant for disagreement, (Lennertz, 2019, 4789). To see why, suppose that a third agent, c, knows that p. While c may respect b's reasoning, she will disagree with b and will believe that the content of b's belief is incorrect, since it is false. Alternatively, even if c does not respect a's reasons for believing p is the case, c does not disagree with a about p. Regardless of how any of the agents formed their beliefs, c will agree with a and disagree with b. The quality of a and b's evidence does not impact c's capacity to agree or disagree with them about p.

This sounds good when we limit our examples to propositional beliefs whose contents are truth-evaluable. True contents are correct, false contents are not, and therefore, agents are correct in believing true contents according to Wedgewood's principle. There is a lacuna in our story, however, should we insist that there are contents that are not truth-evaluable. The task at hand is to provide plausible conditions for content correctness within a dynamic framework that do not appeal exclusively to truth. In addition, these conditions should also explain why Holmes judges the content of Watson's belief to be incorrect.

2.2.2 Candidates for Content Correctness

The literature features a few different accounts of a correct belief that are compatible with the test semantics. While some dynamic accounts appeal to the distinction between content correctness and formative correctness, there are not, to my knowledge, any dynamic accounts that explicitly characterize the conditions for these different kinds of correctness.[20] The proceeding section explores extant proposals, and considers whether they are, in fact, accounts of content correctness or formative correctness. After considering each, I explain why each account is, unfortunately, unsatisfactory.

Evidential Correctness

Willer (2013) deploys the test semantics to engage the puzzle of modal disagreement. Willer's strategy involves adding additional complexity to the semantics described above in supervaluationist fashion in order to capture how epistemic modals express attention towards a possibility. The result is a sophisticated development of the test semantics that addresses a battery of problems involving epistemic modals. After proposing the semantics, Willer addresses an objection against non truth-conditional accounts of epistemic modals from MacFarlane (2011). MacFarlane argues that such accounts are unable to explain the incorrectness judgments which motivate disagreements and retractions, (MacFarlane, 2011, 158). Willer recognizes the importance of correctness judgements, but rejects the claim that no account of correctness is available for the test semantics. He does this by providing his own. Call this condition *evidential correctness*.

Evidential Correctness: A belief is correct if and only if it would not commit someone with the believing agent's evidence to factual error, (Willer, 2013, 64).

According to evidential correctness, an agent is correct to believe some content just

[20]Lennertz (2019), for instance, does appeal to the difference between being wrong to believe and being wrong in believing, but it does not provide an account of precisely what it is to be wrong in believing within the test framework.

in case no one equipped with the same evidence as the believing agent would be committed to factual error. Willer also provides a definition which explains what "factual error" amounts to.

Definition 10: Truthfulness

An information state s is truthful $\Leftrightarrow @ \in s$

An information state is truthful just in case the state has not ruled out the actual world, (Willer, 2013, 56). The actual world is the privileged possible world where all of the propositions that are true at that world, are, in fact, true. An information state that contains any information which is false will rule out the actual world. Thus, an agent is committed to factual error just in case their information state is not truthful.

Evidential correctness is compelling and appropriately broad. It is somewhat intuitive that a "might" belief is correct when the prejacent is compatible with the evidence. Thus, in many cases, it does not attribute error to those who refrain from holding a fully-fledged propositional belief when the evidence does not support that belief. Further, when an agent's evidence does support some belief, φ, and the agent fails to make the appropriate inference and still believes Might $\neg\varphi$, evidential correctness rightly predicts that the believing agent will be incorrect. This is because a more discerning agent with the same evidence would have concluded φ, and thus would be committed to factual error if they believed Might $\neg\varphi$.[21] In addition, evidential correctness applies to propositional and non-propositional beliefs alike. If the proposition associated with a belief is false, then it directly leads to factual error. In addition, certain epistemic modal beliefs that do not express propositions, e.g., "It must be that p" can lead to factual error and can be incorrect as a result. Lastly, while evidential correctness does not appeal directly to truth, the truth remains relevant in that the agent's beliefs should not conflict with the facts. Thus, evidential correctness remains suitably grounded in reality.

[21] The only state that supports both φ and Might $\neg\varphi$ is $\varnothing$, which is not truthful.

Prima facie, evidential correctness appears to make the right predictions in the case involving Holmes and Watson. Suppose that Holmes and Watson have just jointly investigated the same crime scene and have the same evidence. Holmes, in classic fashion, has made inferences that Watson has missed, despite having access to the very same evidence. Based on this evidence, Holmes has miraculously deduced that Moriarty is not the murderer. Were he to maintain these beliefs while also believing the content of Watson's belief, he would be committed to factual error, since $\neg m$ is inconsistent$_D$ with Might m. Thus, the counterfactual conditions for evidential correctness do not hold for Watson's belief, since there is someone, namely Holmes, who has Watson's evidence, yet would be committed to factual error if he held Watson's belief. If Holmes adopted Watson's belief, he would be in the absurd state, guaranteeing that the actual world was removed from his state, and thereby committing him to factual error. Watson is thereby incorrect. Equally importantly, he is incorrect in a way that is accessible to Holmes, allowing Holmes to judge that the content of Watson's belief is wrong and that Watson is wrong in believing that Might m.

There are, however, two substantive problems with this account. The first is that evidential correctness fails to make the right predictions about Holmes' normative judgments in cases where Holmes does not share Watson's evidence. The second is that evidential correctness is sensitive to evidence, and what inferences agents would make based on that evidence. This means not only that evidential correctness can vary with respect to inferential norms, but that it does not concern the correctness of content directly and is instead better characterized as a condition for formative correctness.

How these problems manifest, and that they are, in fact, problems, may not be immediately obvious. With respect to the first, we can see that so long as Holmes shares the same evidence as Watson, he will judge Watson's belief to be incorrect

because Watson has failed to make the requisite inferences based on that evidence. While it is certainly literarily appropriate, given the case at hand, to focus on situations where interlocutors share evidence, evidential correctness fails to make the requisite predictions outside of such carefully engineered scenarios. In particular, it fails in a variety of cases where agents do not share the same evidence. For instance, suppose that the disagreement between Holmes and Watson takes place after the two have each independently investigated separate sections of the crime scene. Holmes' evidence supports $\neg m$ while Watson's provides no evidence for or against m. Each are unaware of the other's evidence and do not share the same evidence. According to evidential correctness, Watson's belief is incorrect just in case someone with Watson's evidence would be committed to factual error, and this is supposed to explain why Holmes can utter, "No, you're wrong," if Watson asserts Might m. The problem here is that Holmes is entirely unaware of Watson's evidence, and is thus unable to judge whether someone with Watson's evidence would be committed to factual error. Nonetheless, he still disagrees, and is perfectly licensed to judge Watson's belief to be incorrect. The problem is not merely one of epistemic access. Suppose, further, that Holmes can read minds, and is aware of Watson's evidence. Even if Holmes can read Watson's mind, if his judgments are guided by evidential correctness, Holmes should conclude that Watson's belief that Might$m$ is correct. Since Watson's evidence is silent with respect to m, no one with Watson's evidence would be committed to factual error. Nonetheless, Holmes still disagrees and can still felicitously utter, "No, you're wrong." Thus, evidential correctness does not make the requisite predictions with respect to the normative judgements required for the disagreement. If Holmes believes $\neg m$, he will judge any Might m belief to be incorrect, whether he is aware of the evidence or not.

If Holmes' correctness judgments are guided by evidential correctness, he will not conclude that Watson's belief is incorrect. This is problematic, since Holmes

is supposed to judge Watson's belief to be incorrect in cases where he disagrees. Recall further that disagreement concerning states of information does not require that conversation manifest, and the absence of conversation does not prevent agents from disagreeing. Thus, Holmes need not even speak with Watson, nor have any ideas about Watson's evidence, in order to disagree. Just like b and c above, Holmes disagrees with Watson because he judges some content to be incorrect, and he is able to do this regardless of the evidence that supports that content. Evidential correctness fails to explain this, and thus fails to motivate the disagreement challenge in cases where disagreeing agents do not share evidence. Things are made worse by the fact that scenarios where disagreeing agents do not have access to one another's evidence are common. In fact, we often engage in active information exchange precisely because we do not share the same evidence. One may even suspect that disagreement between agents that do not share evidence are more common than ones where they do. At the very least, cases where evidential correctness fails aren't recherché, nor need they involve complexities like eavesdropping. Such cases are common if not the norm.

That this issue arises is unsurprising upon consideration of the second problem. Under Willer's account, whether or not some belief is evidentially correct is determined by the evidence, and what inferences would be made by other agents in possession of that evidence. This suggests that evidential correctness is best viewed as a correctness condition for formative correctness rather than content correctness.[22] Relevantly, Willer says the following: "Correctness as characterized above turns on an individual's evidential situation and thus we leave room for the possibility that, given adequate variation in what is known, [Watson] correctly believes that [Moriarty might be the murderer], while [Holmes] correctly believes that [Moriarty is not the murderer]... But we are well advised to bear in mind that correctness thus under-

[22]Willer (2013) discusses correctness, but does not distinguish content content from formative correctness. I suspect that some of the problems associated with evidential correctness may be due to a conflation of the two.

stood cannot serve as a reliable guide to the semantics and pragmatics of epistemic modals," (Willer, 2013, 64-65).[23] I agree with Willer's assessment here. This is not the kind of correctness that should inform our semantic theorizing. The problem is that the disagreement challenge that motivates adoption of the test semantics in the first place is, itself, motivated by Holmes' correctness judgment. This is to say that the kind of correctness we are searching for is precisely the kind that *does* motivate our semantic theorizing. Should we trivialize Holmes' judgment, the contextualist is free to trivialize it as well, and the disagreement challenge dissolves. Thus, while evidential correctness may well be a plausible constraint on formative correctness for belief, it cannot play the dialectical role of the theory of correctness we are after.

What we need is a correctness condition for dynamic contents that does inform our semantic theorizing, while also vindicating the requisite correctness judgments that motivate the disagreement challenge. Two options readily suggest themselves.

Weak Correctness

With respect to content correctness, an agent's ability to be correct in believing some content depends entirely on the semantic value of that content. Not so for formative correctness, where an agent can be perfectly formatively correct, even if the content is not. This suggests that conditions for content correctness should concern content directly and should not appeal to evidence or inferential mechanisms. This simultaneously explains why evidential correctness failed, while pointing us in a more productive direction. I next consider a proposal in this spirit, which I call *weak correctness*.

Weak Correctness: If an agent a holds a belief with content φ the content of that belief is correct if and only if $@ \in I_a[\varphi]$.[24]

[23] I alter Willer's quote by exchanging Holmes and Watson with the names of Willer's example subjects, Alex and Mary.

[24] Here and henceforth, I characterize contents in terms of formulas as opposed to updates, but

Weak correctness holds that a belief is correct if and only if that belief does not commit the believing agent to factual error. This is to say that a belief is correct just in case that belief does not cause the believing agent to believe falsely.

Weak correctness has some immediate advantages. The first is that it does not appeal to evidence or inference, ensuring that it characterizes correctness of content. The second is that, like evidential correctness, it applies to propositional and non-propositional beliefs alike. If the content of some proposition is true, then it will be correct. If it is false, it will be incorrect. Things change with respect to epistemic modal beliefs, which are neither true nor false. Consider the content Might p. A belief in this content will be correct if the believing agent's information state contains at least one p world. It will be incorrect if it does not contain any p worlds. More generally, we see that some epistemically modalized content is correct just in case the test performed by the modal is passed. This both explains why it is incorrect to believe Might p when one believes $\neg p$, but it also explains the intuition that "might" beliefs can be correct when an agent is undecided as to the truth of some proposition.

The obvious problem, however, is that weak correctness universally fails to vindicate Holmes' judgment that Watson's belief is incorrect. Since Watson is undecided, his belief that Moriarty might be the murderer does not commit him to factual error, and is thereby correct. Moreover, even though Holmes believes $\neg m$, Holmes will recognize that Watson's belief that Might m will not commit him to factual error. Thus, Holmes should judge Watson to be correct in believing Might m. This problem generalizes to all instances of modal disagreement and universally fails to predict the incorrectness judgments that motivate the problem. Thus, whatever its advantages, weak correctness cannot be the correctness condition that motivates the disagreement challenge.

---

this choice is purely practical, and nothing hangs on it. Definitions concerning content can be easily altered to instead appeal to updates rather than formula.

Strong Correctness

Willer (2013) mentions an alternative in a footnote on pp.63, which appeals to a proposal from Yalcin (2011). Yalcin proffers a condition for correct assertion called *advisability*, but we can instead articulate this as a correctness condition for belief and explore the results. Willer does not pursue this condition in depth, but, in light of my objections, we can take a closer look.

Strong Correctness: A belief is correct if and only if an agent who knows all of the relevant facts holds this belief, (Willer, 2013, 63).[25]

According to strong correctness, a belief is correct just in case some idealized agent who already knows the relevant facts holds this belief. Thus, if Holmes and Watson are debating about m, the idealized knower will know whether or not m is the case. Strong correctness is significantly stingier than weak correctness when it comes to whose beliefs are correct, specifically with respect to "might" beliefs. Just like weak correctness, any commitment to factual error will result in a belief that is incorrect. However, weak correctness was permissive with respect to some epistemic modal beliefs. Even if some proposition p was false, weak correctness allowed for beliefs like Might p to be correct, so long as they did not commit the believer to factual error. Not so for strong correctness. According to strong correctness, if p is false, then an agent who knows the relevant facts will not believe Might p, and so, any such belief is incorrect. Thus, it is not enough to merely fail to believe falsely. In order to satisfy strong correctness, one must believe truly.

We can now apply strong correctness to the disagreement between Holmes and Watson. Holmes believes $\neg m$, which is to say that Holmes believes the facts are such that an idealized agent would believe $\neg m$, and would not believe Might m. This

[25]This definition appeals to a locally idealized knower with access to the relevant facts. Those with concerns about the work that relevance plays may appeal to an absolutely idealized knower who knows all of the facts. What follows is equally applicable to both characterizations, so the distinction is not emphasized.

explains why Holmes judges the content of Watson's belief to be incorrect. Note that this applies even if it turns out that Holmes himself is mistaken, and Moriarty is in fact the murderer. In virtue of believing $\neg m$, Holmes believes that the facts are such that $\neg m$ and subsequently believes that an idealized agent would also believe $\neg m$.[26] This is looking good, since it explains Holmes' judgment that Might m is incorrect. Consequently, this predicts that Holmes will judge any agent who believes Might m to be incorrect in believing this content. One can change the scenario as freely as one likes, adding differing evidence, eavesdroppers, or whatever else. Strong correctness predicts that Holmes will judge any belief that Might m to be incorrect, as is required to motivate the puzzle.

Unlike weak correctness, strong correctness predicts that Holmes judges the content of Watson's belief to be incorrect, maintaining the initial motivation for modal disagreement. Unlike evidential correctness, strong correctness concerns content exclusively, and is thus an appropriate condition on correctness of content. In addition, it allows Holmes to make the requisite correctness judgment about the content of Watson's belief in all cases, regardless of evidence. Strong correctness thus avoids the challenges faced by evidential and weak correctness and appears to be the most attractive option thus far. Unfortunately, it also makes a multitude of less attractive predictions concerning individuals who are undecided. Recall that for the test semantics, to be undecided about some proposition p is to have an information state that can sustain update with p as well as $\neg p$. Unlike traditional propositional accounts, where indecision is the absence of belief, our current framework ensures that indecision is captured by a belief. If an agent is undecided with respect to p, then their information state will support Might $p \wedge$ Might $\neg p$.[27] It doesn't matter whether this

[26]Put another way, when it comes to our propositional beliefs, we tend to think of ourselves as locally idealized agents, since we think our own propositional beliefs to be true. We can be mistaken, however, about what an idealized agent would believe. This also aligns nicely with the observation made by Willer and others that agents tend to test "might" beliefs "against their own perspective," (Willer, 2013, 87).

[27]The same agent will also believe Might p and Might$\neg p$.

beliefs are occurrent, or whether the agent recognizes that they have it. Importantly, it is not that indecision yields, produces, or causes this belief. To be undecided about p *just is* to believe Might $p \wedge$ Might $\neg p$.

In a purely propositional framework, strong correctness appears plausible and attractive.[28] However, in frameworks where indecision has content, there are some rather extreme consequences.

Ubiquity of Error: For any propositional content φ, if $@ \in [\![\varphi]\!]$, then any doxastic agent who fails to believe φ, and thereby believes Might $\neg\varphi$ believes some content that is incorrect and thus, is wrong in believing Might $\neg\varphi$.

According to strong correctness, any agent who fails to believe truly with respect to any proposition is guaranteed to be incorrect. This includes both agents who believe falsely, as well as agents who are undecided. A immediate consequence is that every actual doxastic agent is incorrect. Both weak and evidential correctness missed crucial cases of incorrectness and thus failed to motivate the disagreement challenge. Strong correctness avoids this problem by painting with an especially broad brush and characterizing any content that is not consistent with the facts as incorrect. It doesn't matter whether the inconsistency is the product of indecision or false belief.

It is important to appreciate that the traditional, static framework operates differently. According to Wedgewood's truth norm, the content of a belief is correct just in case it is true. False beliefs are incorrect, but the absence of belief is neither correct nor incorrect with respect to the truth norm. Classically, to be undecided with respect to some proposition is to not have a belief towards that proposition. While indecision may be correct in some circumstances and not others, this correctness is not governed by any content norm–there is no content–and is instead governed by some formative norm of belief. As a result, the classical framework does not pre-

[28]If we interpret strong correctness as suggested in footnote 26, then it is equivalent to Wedgewood's truth norm in frameworks that are propositional.

dict such large-scale error in content. Unlike classical frameworks, the test semantics treats indecision itself as having content. When coupled with strong correctness, the test semantics is able to explain how the content of certain "might" beliefs are wrong. The consequence, it turns out, is that a whole lot of content, including all indecisive content, is wrong.

This might sound like a bullet worth biting. After all, false content and indecisive content each fail to accurately represent the world. In addition, it is intuitively plausible that content that is inconsistent with what one believes should be judged as incorrect. Moreover, it is worth repeating that strong correctness makes no judgments about what is correct to believe. Strong correctness only governs what one is correct in believing. Watson is most plausibly correct to believe $\neg m$. Strong correctness merely predicts that he is incorrect in believing it. Unfortunately, other consequences of strong correctness are not so easily embraced.

Homogeneity of Error: There is no normative difference between indecisive content and false content. Both are incorrect in exactly the same way, for exactly the same reason.

The homogeneity of error arises from the fact that strong correctness does not discriminate between different kinds or degrees of content correctness. Believing indecisive content is just as incorrect as believing falsely. Thus, there is no difference in correctness between being merely ignorant about the world and misrepresenting the world.

The test semantics recognizes contents that express indecision directly, without describing the agent as being in an undecided state. Accordingly, correctness conditions for such contents seem obliged to distinguish between the manner in which one is incorrect in being ignorant, and the manner in which one is incorrect in believing falsely. Strong correctness fails to do this. What's more is that this distinction is clearly suggested by the test semantics itself. To see this, we can enrich the scenario

with Holmes and Watson. Suppose that the interaction also includes a third agent, Lestrade, who believes that Moriarty is the murderer (m). Suppose further, for objectivity, that $@(m) = 0$. Observe that while both Watson and Lestrade are, in some sense, incorrect, they are in radically different doxastic situations. Watson, in virtue of being undecided, can still consistently come to know the truth that Moriarty is not the murderer. Watson believes Might $m \wedge$ Might $\neg m$, however, the sequence, $\langle$Might $m \wedge$ Might $\neg m, \neg m\rangle$ is dynamically consistent.[29] Thus, the contents of Watson's beliefs do not impede him from arriving at the truth. Lestrade's situation is different. Lestrade believes m, which is dynamically inconsistent with the factual update expressed by $\neg m$. Thus, Lestrade's false belief not only ensures that his information state will not be truthful, but it actively prevents him from productively updating with the truth. Thus, Lestrade will need to reject his belief that m if he is to have any hope of arriving at the truth. It should be clear that, based on the semantics alone, Lestrade, who believes falsely, is incorrect in a way that Watson is not. Thus, we observe a significant discrepancy. The test semantics and, hopefully, our intuitions suggest a normative distinction between indecisive content and false content. Strong correctness, the content governing norm responsible for making such individuations, is not sensitive to this distinction, and treats indecisive content as wrong in the same way as content that is false.[30] All doxastic sins are weighed in equal measure.

One may again be tempted to take the homogeneity of error in stride. However, if strong correctness is to play the role we need it to, there are further, even more extreme consequences. Recall that our principle must be such that it guides the normative judgments of doxastic agents. In our case, it needs to guide Holmes to conclude that the content of Watson's belief is incorrect. While strong correctness

[29]This is because many Might updates are not *persistent*.

[30]Data gathered in Khoo (2015) suggests that speakers are sensitive to this distinction, or one very much like it, insofar as experimental subjects are inclined to judge "might" beliefs with false prejacents as wrong, but not false.

gets the judgments right with respect to Holmes, it suggests that others, including Watson, should makes some dubious judgments.

Generalized Error Attribution: For any agents a and b, if b believes that a is undecided about some proposition p, then b must judge the content of a's belief that Might $p \wedge$ Might $\neg p$ to be incorrect.

Our story about correctness principles begins with Holmes' judgment that Watson's belief is incorrect. Strong correctness provides a satisfying story about not only the correctness/incorrectness of Watson's belief, but, also why Holmes makes the judgment that he does. In order for this story to work, our correctness principles can't merely exist in the background, but must, on some level, guide the judgments of agents like Holmes. Such principles, however, are not exclusive to Holmes and should also guide the judgments of other doxastic and conversational agents, including Watson.

Consider any scenario where some agent a knows that some agent b believes Might $p \wedge$ Might $\neg p$. Regardless of whether a believes whether p, a knows that p must either be true or false, and thus knows that an idealized agent equipped with the facts would believe either p or $\neg p$, and, in virtue of this, would not believe Might $p \wedge$ Might $\neg p$. Any state that supports either p or $\neg p$ would crash upon update with Might $p \wedge$ Might $\neg p$, and thus, the content of b's indecisive belief will be judged as incorrect by a.

Thus, strong correctness predicts that agents should always judge indecisive content as wrong. This is not what we observe.

(37) a. Poirot: I have no idea who the murderer is.

b. Hastings: John had a lot to gain from Inglethorpe's death. He might be the murderer, but he might not be.

c. Poirot: # That's wrong.

It is strange for Poirot to assert that the content of Hastings' undecided belief is incorrect. This is, at least in part, because Poirot himself is undecided. Strong correctness predicts, surprisingly, that this is precisely what should happen. In practice we observe the opposite: undecided agents do not appear to disagree with other undecided agents, and it strikes as infelicitous if they judge indecisive contents to be incorrect. This prediction does not seem to model the way speakers use normative judgments in conversation. Thus, strong correctness overpredicts error attribution.

Perhaps more interestingly, generalized error attribution has introspective cases as well. In many circumstances, undecided agents are aware that they are undecided with respect to a particular issue. Call this *conscious indecision*. Indeed, this seems to be Watson's situation with respect to m. Watson does not know whether m is true or false, but he does know that an idealized agent would believe either m or $\neg m$, and would not believe $\text{Might}\, m \wedge \text{Might}\, \neg m$. Thus, the content of Watson's indecisive belief is incorrect, and *Watson knows it*. He should thus judge his own belief to be incorrect. Strong correctness was attractive insofar as it made this judgment available to Holmes. The problem is that it makes the judgment equally available to Watson. If this is the case, then we would expect utterances like the following:

(38) # Moriarty might be the murderer, and he might not be, but this is wrong.

(3.2) sounds horrible and certainly isn't how people articulate indecision. It not only sounds bad but has an air of contradiction about it. Notice that the infelicity projects when embedded.

(39) # Suppose that Moriarty might be the murderer, and that he might not be, and that this is wrong.

(40) # Mycroft believes that Moriarty might be the murderer, and that he might not be, and that this is wrong.[31]

[31]Infelicity also projects when embedded under conditionals, but the sentences are incredibly clunky: "If Moriarty might be the murderer, and he might not be, and this is incorrect, then..."

The projected infelicity of (3.2) suggests the problem is not merely one of assertion, but that the content of (3.2) is paradoxical or contradictory. Indeed, it seems to be something like the normative corollary of the liar's paradox. Instead of describing previous portions of the sentence as a lie or false, it characterizes them as incorrect, with similar results. Despite the contradictory air of these sentences, strong correctness predicts these are the judgments that consciously undecided agents should make. This is a puzzling result. Before strong correctness can be considered tenable, its proponents will need to resolve the putatively paradoxical nature of of reflexive error attributions, in addition to explaining why undecided agents do not express negative judgment about the beliefs of other undecided agents.

The final problem arises when we consider how normative judgments motivate our actions. Upon realization that some content is incorrect, agents should cease to believe that content, and revise their beliefs accordingly. We also expect this from our interlocutors. Holmes does not make a normative judgment about Watson's belief with the hope that Watson will maintain it. Rather, Holmes is trying to convince Watson to believe $\neg m$. If indecisive content is just as wrong as false content, and we recognize our own indecisive belief content to be wrong, we should react to our own indecisive beliefs in the same way that we react when we realize that our beliefs are false, by rejecting them.

Forced Decision: An agent who judges the content of an indecisive belief to be incorrect should reject the content of that belief.

Generalized error attribution has it that any consciously undecided agent should judge the content of their indecisive belief as wrong. If they reject contents that are wrong, then they should reject indecisive contents. This means that if an agent is consciously undecided about the truth of some proposition, then that agent should reject their indecisive belief and, thereby, decide. Put simply, whenever an agent knows that they are undecided, they should guess.

Consider a straightforward case of false belief. Suppose a believes that p. Later, a discovers that p is false. They thus judge their belief that p to be incorrect. In response to this, they revise their beliefs to state such that $s_a \vDash \neg p$. So far so good, but strong correctness treats indecisive contents in exactly the same way. If a consciously believes Might $p \wedge$ Might $\neg p$, then they should judge it as incorrect, and thereby revise their beliefs such that $s_a \vDash \neg(\text{Might } p \wedge \text{Might } \neg p)$. But the only states that support $\neg(\text{Might } p \wedge \text{Might } \neg p)$ are states that support p and states that support $\neg p$. The only state supporting both is the absurd state, $\varnothing$. Thus, a is forced guess. What is especially interesting is that, as far as strong correctness is concerned, there is no risk in this guess. Any state that supports $(\text{Might } p \wedge \text{Might } \neg p)$ is guaranteed to be incorrect. However, by guessing, the agent is only increasing their chances of being correct. If they guess wrong, they are no worse off than if they stayed undecided. If they guess right, they have disabused themselves of believing incorrectly.[32] Accordingly, should we adopt strong correctness for the test semantics, we would expect undecided agents to guess in an attempt to avoid error. This is neither what we observe in everyday thought and talk nor what we should want from rational agents.

By my lights, things look fairly dire for strong correctness. The idea that every undecided agent is incorrect in being undecided is a bitter, albeit manageable, pill to swallow. However, the idea that such agents should also judge all indecisive content as incorrect, and guess whenever they are undecided, is simply too much to choke

[32]We might be inclined to separate two normative notions for doxastic action: one that drives us to believe correctly, and one that deters us from believing incorrectly. We may further place these norms in a hierarchy where one trumps the other. For example, If we prefer correct beliefs more strongly than we dispref er incorrect ones, one would be disposed to guess, perhaps in runaway fashion, because this would yield a net positive. To avoid rampant guessing, we may instead choose to flip the order of the hierarchy, and make incorrect beliefs more undesirable than correct beliefs are desirable. One would expect this to prevent rampant guessing, since the negative cost of getting it wrong outweighs the positive cost of getting it right. In a propositional framework, this is exactly what you'll get, making the more conservative hierarchy of the preferences the better option. This fails within the current framework, however, since it is the pressure to avoid incorrectness which motivates guessing. Thus, within the current framework *either ordering of the norms will yield a guess*, so hierarchies like these will not avoid the puzzle.

down. In its current form, strong correctness appears untenable.

Where to Go from here

This section has argued that evidential correctness, weak correctness, and strong correctness are each untenable. Evidential correctness and weak correctness failed to yield the correctness judgments that motivate the disagreement challenge. Strong correctness succeeded in this regard, but yielded doxastically implausible consequences. The problems with strong correctness seem to stem from the fact that indecision has content within the test semantics. This suggests one of two strategies to avoid the problem. The first is to separate believing Might $p \wedge$ Might $\neg p$ from indecision. This option is explored, with different motivations, in Willer (2013) and Yalcin (2018). While each of these options will mitigate certain instances of the problems, they continue to persist. I address each of these strategies briefly in the next section. The second, more attractive, option is to recognize that indecisive content is wrong in a different way than false content, and alter our account of correctness accordingly. This option is developed in §4.

2.3 Rethinking Indecision

In an effort to make my argument as general as possible, I have provided a fairly simple version of the test semantics. The literature, however, features more sophisticated proposals, like Willer (2013) and Yalcin (2018), that separate indecision about p from believing Might $p \wedge$ Might $\neg p$. Since my criticisms of strong correctness assume that these coincide, these accounts will have to be addressed.

The version of the test semantics in Willer (2013) is complex and resists brief summary. A crucial feature, however, is that information states have added structure that alters the update effects of epistemic modals. For Willer, information states are

not sets of worlds, but sets of sets of worlds. Call each set of worlds within a state a *substate*. When an update is performed on a state, updates are applied to each substate, and then substates are aggregated. The result is that epistemic modals do not simply test for compatibility with the prejacent. Rather, a state will only support Might p on condition that every substate contains at least one p world. This is intended to capture the idea that sentences like Might p express more than mere compatibility with p, but that p is taken seriously, or matters, in inquiry, (Willer, 2013, 56). Accordingly, an agent's information state will only support Might p if that agent is disposed to take p seriously as a possibility.

A promising result for Willer's semantics is that agents who are not disposed to take both p and $\neg p$ seriously will not believe Might$p \wedge$Might$\neg p$, and can thereby avoid error. This will include many agents who simply never have considered p.[33] The result is that a large portion of undecided agents will no longer be incorrect in believing Might $p \wedge$ Might $\neg p$, since they do not need to believe that in order to undecided. We see a similar result with generalized error attribution and forced decision. Since not all undecided agents will believe Might $p \wedge$ Might $\neg p$, they will not be judged as incorrect, nor will they be pressured to guess.

This interaction with strong correctness is undeniably an improvement over what was previously discussed. However, strong correctness still yields unhappy results when coupled with Willer's proposal. In particular, the account falters in instances where undecided agents take both alternatives seriously. Unfortunately, many of the most important and difficult instances of deliberation are precisely this way. Watson, for example, is aware of Moriarty's past crimes, and takes the possibility that Moriarty is the murderer seriously. However, he is not so foolish as to believe that this is the only possibility worthy of attention and takes the possibility that someone else is

[33]Many, but perhaps not all. For Willer, considering something to be a relevant alternative is a matter of disposition, (Willer, 2013, 50). This, in principle, leaves open the possibility for an agent who has never considered p to be disposed to take p seriously, upon consideration.

the murderer (and Moriarty is not) equally seriously. In such cases, the negative effects of generalized error attribution and forced decision remain, as the content of Watson's belief that Might $m \wedge$ Might $\neg m$ should be judged incorrect, even by Watson himself. Forced decision also looms, although with a new wrinkle. In the simple test semantics, the only way to reject the belief that Might $p \wedge$ Might $\neg p$ was to decide. Willer's account offers a second way to reject the belief: cease to take at least one of the possibilities seriously. The good news is that agents who believe Might $p \wedge$ Might $\neg p$ are no longer forced to guess. The bad news is that the new option isn't much better, as ceasing to take a relevant alternative seriously sounds similarly irrational. One might try to bite the bullet here, and think that this is still a better option. In doing so, one might embrace a weaker principle whereby agents should not consider incompatible alternatives seriously at the same time. If this principle holds, we would expect sentences like the following to be infelicitous:

(41) It might be raining, and it might not be.

However, we felicitously utter sentences like (41) regularly, and the principle strikes as untenable. Thus, the proposal made in Willer (2013) still suffers from critical problems associated with strong correctness.

Yalcin (2018) does not propose a semantics, but rather, an account of belief where belief is *question sensitive*. On this account, beliefs are relativized to questions or *resolutions* of logical space, (Yalcin, 2018, 30). The idea is that agents can fail to recognize certain distinctions in the logical space. When an agent does recognize a distinction, her logical space will represent at a higher resolution, individuating possibilities that it previously did not. The account is intended to solve various puzzles concerning logical omniscience, but, the view allows for a distinction between indecision about p, and believing Might $p \wedge$ Might $\neg p$. For instance, if an agent has never considered the issue of whether or not p, their logical space will not individuate p worlds from $\neg p$ worlds. The same goes for agents that lack some relevant concept

crucial to understanding p, (Yalcin, 2018, 34). The pertinent result, with respect to strong correctness, is that agents whose information state lacks the appropriate resolution for some proposition p, will not believe Might $p \wedge$ Might $\neg p$, despite being undecided. The result is much the same as it was for Willer, in that the ubiquity of error is substantively mitigated. Undecided agents who have never considered whether p will not believe Might $p \wedge$ Might $\neg p$, and thereby will not hold an incorrect belief.

Unfortunately, this strategy is of no help for agents whose logical space does respect the appropriate resolution. Assuming that agents know what they are saying, this will include any agent who asserts a "might" sentence. Thus, agents who ask whether p, but remain undecided, will believe Might $p \wedge$ Might $\neg p$ and run into all of the same problems brought on by strong correctness. There is a similar wrinkle associated with forced decision, where an undecided agent is not forced to decide, and instead, may choose to cease asking the question, and coarsen the resolution of their logical space in the face of the incorrectness of their own ignorance. This result is decidedly unattractive.

Thus, we observe that separating indecision about p from believing Might $p \wedge$ Might $\neg p$ only helps to a limited extent, and does little to solve the hard cases. Relevantly, neither solution appears to help us with Watson and Holmes. It is worth emphasizing that nothing I've said here amounts to rejection or criticism of either view. Rather, I only argue that they do not solve the problems that arise from strong correctness. Indeed, it would appear that any account where an undecided agent *can* believe Might $p \wedge$ Might $\neg p$ will face similar troubles. It is difficult to see how an account would successfully separate *all* instances of indecision from believing Might $p \wedge$ Might $\neg p$, without dramatically altering the test conception of "might". This suggests that the observations about Willer (2013) and Yalcin (2018) generalize, and the issues caused by strong correctness will remain recalcitrant for any plausible

view in this tradition. Thus, it does not appear that our account of belief, or the semantics itself, is the locus of the problem. Rather, strong correctness appears to be the culprit.

2.4 Multiple Content Norms

Evidential, weak, and strong correctness each suffer from different issues, but I suspect that a common cause lies at the root of their respective problems. For truth-conditional frameworks, Wedgewood offered a single content norm in the form of the truth norm. Things like indecision lack content and are thus governed by formative norms, making a single content norm perfectly appropriate. The previously considered proposals each attempt the same strategy, modified to fit the test semantics. This strategy, however, seems destined to fail, as the test semantics admits of different kinds of contents that aim to do very different things. Propositional contents aim to represent the way the world is. Non-propositional contents, whatever they do, don't do that. It is thus little wonder that a single norm is unable to govern such disparate aims. One might also suspect that some of the previous accounts were not beyond repair and could be buttressed with further principles in order to make the right predictions. What follows is an attempt to do right by both of these intuitions.

I propose that adoption of the test semantics should be coupled with the recognition of two content norms. According to these norms, the correctness conditions for every kind of content will not always be the same. Not only will this allow us to distinguish the ways in which different kinds of content can be differently correct or incorrect, but it will also allow us to integrate features of aforementioned views that were attractive. In particular, the proposal appeals to a condition inspired by weak correctness, as well as a second principle which closely mirrors strong correctness. The difference is that neither is implemented in a vacuum, as each are smaller parts

of a larger account. Thus, proponents of the previously considered views may wish to view my proposal as a development of these accounts rather than a direct competitor.

2.4.1 Locative Correctness and Informational Correctness

Our goal is to draw a distinction between two kinds of content and characterize norms which should govern each. An initially tempting strategy might be to say that propositional contents are governed by something like the truth norm, and non-propositional contents are governed by another. Unfortunately, this won't work. Some non-propositional contents like Must$p$ and $p \wedge$ Might$q$, have propositional entailments, and are, in some sense, beholden to norms sensitive to truth, despite the fact they they are neither true nor false. This suggests that at least some non-propositional contents need to be governed by the same norms as propositional ones.

Given that some non-propositional contents have propositional entailments, we might instead draw our distinction by individuating contents by their propositional entailments. We could do this by saying that some content is correct when its propositional entailments are true.[34] This won't work either, since all contents, propositional and otherwise, entail tautologies, which are propositional. Thus, things like Might p will always be correct, since all of their propositional entailments are tautologies. While this strategy doesn't work, it suggests that tautological entailments are the problem. We can avoid these by appeal to *locativity*.

Definition 11: Locativity

φ is locative $\Leftrightarrow$ there exists some contingent proposition $[\![\psi]\!]$ s.t. $\varphi \Vdash_D \psi$

Some content is locative if and only if that content dynamically entails some propositional content ψ, where $[\![\psi]\!]$ is contingent.[35] Locative content individuates certain possibilities from others. Paradigmatically, contingent propositions do just this. They

[34]This condition is actually equivalent to weak correctness.
[35]A proposition $[\![\varphi]\!]$ is *contingent* $\Leftrightarrow [\![\varphi]\!] \neq W$ and $[\![\varphi]\!] \neq \varnothing$.

are true at some possible worlds, and false at others. To believe a contingent proposition is to believe that the actual world is a particular way and is not some other way. Similarly, to believe some contingent propositional content is to locate oneself non-trivially in the logical space.

Certain instances of locativity are obvious. Atomic sentences are locative since for any atomic sentence $\mathrm{p}, \mathrm{p} \Vdash_D \mathrm{p}$. The same holds for contingent propositional contents. Alternatively, "might" sentences where the prejacent expresses a contingent proposition are not locative. Sentences like Might p do not entail any contingent proposition and thus, do not individuate possibilities from others. Other cases may be less obvious. All contradictions are locative since they yield the absurd state, which accepts every formula, contingent propositions included. Tautologies are not locative since they are accepted by every state and do not entail any contingent proposition. This makes sense when we consider that tautologies do not individuate any world form any other, and thereby, do not locate in any non-trivial sense. Lastly, "Must" sentences where the prejacent expresses a contingent proposition are locative. Must p, for instance, is locative, since Must $p \Vdash_D p$. The same holds for "mixed" conjunctions like $p \wedge$ Might q, since $p \wedge$ Might $q \Vdash_D p$.

Locative content does not coincide with content that is propositional. Many non-propositional contents, most notably Must p, are locative. Similarly, some propositional contents are not locative, e.g., tautologies. Locativity is intended to distinguish those contents which come bundled with commitments about how the world is from those that do not. Once we distinguish locative content from non-locative content, we can characterize how locative content can be correct. Some locative content will be correct when its propositional consequences accurately represent the world. More rigorously:

Locative Correctness: Some locative content φ is locatively correct if and only if for every contingent proposition $[\![\psi]\!]$, if $\varphi \Vdash_D \psi$ then $@ \in [\![\psi]\!]$.

Locative content is locatively correct just in case every contingent proposition entailed by that content is true at the actual world. Somewhat loosely, content is correct only when it supports the facts. Locative contents will be locatively correct when they locate the believing agent in the same portion of logical space that contains the actual world. We can project locative correctness to the agential level as well.

Agentially Locative Correctness: An agent, a is locatively correct in believing some locative content φ if and only if φ is locatively correct.[36]

Agents will be locatively correct in believing locative contents just in case the contingent propositional commitments incurred by belief in those contents are true. This ensures that correctness at the agential level is determined purely by the content of the belief. It also ensures that locative correctness remains a content norm, and does not encroach into formative territory.

Locative correctness makes the expected and desired prediction about locative contents. Propositional contents that are true (excluding tautologies) are locatively correct. When they are false, they are incorrect. A similar story goes for locative beliefs like Must p, believing which will be locatively correct when p is true at the actual world. If p is false, it is incorrect. All contradictions are locatively incorrect, as they should be. Agents who hold any of these locative beliefs inherit their normative status and are similarly locatively correct/incorrect.

Locative correctness can be seen as something of an extension of the truth norm of belief, but, it accommodates non-truth apt content that comes bundled with truth-apt commitments. Thus, locative belief aims at truth, albeit sometimes indirectly. Rather than requiring the content of the belief to be true, it requires that its nontrivial propositional consequences be true. Moreover, it captures the ideal of representational belief: to accurately represent the world. It does not, however, tell us anything

[36]I define correctness at the agential level and the content level. One may also, or instead, choose to define it at the belief level. The difference is trivial since correctness is ultimately decided by the content.

about contents or beliefs that are not locative. Locative correctness applies only to locative contents and, derivatively, locative beliefs, so all non-locative contents are neither locatively correct, nor locatively incorrect. This makes sense, since they do not attempt to individuate certain possibilities from others. Instead, I propose that non-locative contents are held to a slightly different standard. Call this *informational correctness*.

Informational Correctness: Some content φ is informationally correct if and only if $\{@\} \vDash \varphi$.

Some content is informationally correct just in case an idealized state that contains only the actual world supports that content. More intuitively, content is informationally correct if it fits with the facts. Here, we see strong correctness in a maximally idealized form.[37] The important difference is that informational correctness does not have to work alone, and, as we shall see, interacts with locative correctness in a way that avoids the problems of strong correctness.

Informational correctness provides correctness criteria for all contents, including non-locative contents. For example, if p is true at the actual world, then Might p is informationally correct. Given the same facts, Might $\neg p$ is informationally incorrect. Tautologies are vindicated since all tautologies are informationally correct. Informational correctness applies to locative contents as well. For instance, non-tautological propositional contents that express true propositions, are informationally correct. False ones are informationally incorrect. Epistemically modalized contents like Must$p$, where p is true, are informationally correct, and when the p is false, incorrect. Contradictions are universally informationally incorrect. Importantly, contents that are locatively incorrect are guaranteed to be informationally incorrect, but not *vice versa*, meaning that locative incorrectness is a special case of informational incorrectness.[38]

[37] This formulation is equivalent to the one mentioned in footnote 24.

[38] Interestingly, locative correctness is not a special case of informational correctness. Suppose that $@ \in [\![p]\!]$ and $@ \notin [\![q]\!]$. $p \wedge$ Might q is locatively correct, but informationally incorrect.

More on this momentarily.

Like locative correctness, informational correctness also projects to the agential level.

Agential Informational Correctness: An agent, a is informationally correct in believing some content φ if and only if φ is informationally correct.

Notice that informational correctness also indirectly concerns truth, but in a different way than locative correctness. Locative correctness demands that any truth-apt entailments of a given locative content be true. Informational correctness instead demands that contents with or without truth-apt entailments 'fit with' the facts. In this sense, informational correctness guides inquiry towards the truth, without representing it. Thus, the truth norm remains in spirit, while not in the letter. We also see echoes of weak and strong correctness. Locative correctness is inspired by weak correctness, but only concerns contingent consequences. Informational correctness is an idealized form of strong correctness. Unlike before, the two principles interact in a way that allows for finer grained distinctions than were previously available.

To believe some locatively incorrect content is to rule out possibilities that one should not. Thus, we associate locative incorrectness with false contents and false beliefs. Any agent who holds a locatively incorrect belief will not have an information state that is truthful. On the other hand, some content is informationally incorrect when it leaves open possibilities that it ideally would not. However, one can hold a belief that is informationally incorrect while still having an information state that is truthful. Thus, we associate informational incorrectness with ignorance or indecision. This is to say that to be locatively incorrect in believing is believing falsely, while being informationally incorrect is failing to believe truly. Believing falsely is a special case of failing to believe truly, so agents that are locatively incorrect are also informationally incorrect.

I take this to be an extremely natural and plausible distinction, and it isn't ex-

actly new. Traditionally, however, the normative distinction between ignorance and misrepresentation crossed the boundary between content norms and formative norms. Since the test semantics characterizes both false belief and indecision at the content level, we should expect two content norms in order to distinguish the normative status of each. Indeed, semantics that deviate from canonical truth-conditional frameworks are often motivated by the intuition that we ought not to treat contents that express uncertainty in the same was as we treat truth-apt contents. Those that find this attractive should have little trouble extending this reasoning to the norms that govern content as well.

2.4.2 Problems Solved

Recognition of locative and informational correctness avoids the second objection to evidential correctness, since neither norm concerns evidence nor strays into formative territory. The proposal avoids other problems as well. The ubiquity of error is now feature rather than bug. Large scale incorrectness is predicted, but in an appropriately fine grained way. Agents who believe falsely are locatively incorrect, as they should be. Agents who are ignorant or undecided, are also incorrect, but merely informationally so. Insofar as we associate mere informational incorrectness with ignorance, this is exactly right. Relatedly, the homogeneity of error no longer holds, since undecided agents are incorrect in a manner that is distinct from agents who believe falsely.

We turn again to Holmes and Watson. Holmes believes $\neg m$ and, therefore, believes that m is false at the actual world. He thus judges Watson's belief that $\mathrm{Might}\, m$ to be informationally incorrect. When Holmes says "No, that's wrong," we consider this a judgment of informational incorrectness. More generally, we associate the incorrectness judgments made by disagreeing agents with judgments of informational incorrectness. This is applicable to modal disagreements, as well as straightforward instances of disagreement where some agent holds a propositional belief, and an-

other believes its negation. We can see this by again enriching the scenario to include Lestrade, who believes m. Holmes disagrees with both parties, and is inclined to think that both Watson and Lestrade are informationally incorrect. However, Holmes takes Watson to be merely informationally incorrect, while he believes Lestrade to be locatively, and informationally incorrect. Consequently, Holmes need not believe that Watson's information state fails to be truthful, but he must believe that Lestrade's does. This distinction vindicates the intuition that Holmes thinks that Lestrade wrong in a way that Watson is not. This is to say that Holmes recognizes a normative difference between Lestrade's and Watson's doxastic positions, despite the fact that they are both informationally incorrect. Importantly, this is a distinction that non-expert interlocutors are sensitive to. This is supported by data gathered in Khoo (2015), which observes that in instances of modal disagreements, subjects regularly recognize that agents like Holmes need not believe the content of agents like Watson's belief to be false in order to disagree, (Khoo, 2015, 520). This suggests that participants can distinguish beliefs that misrepresent the world (and are false) from those that do not represent the world at all but can still be rejected. The difference between locative correctness and informational correctness explains this distinction in addition to vindicating the correctness judgments that motivate the disagreement challenge. The example generalizes, and the same explanations are available across contexts, including eavesdropper cases.

Individuating informational and locative correctness enables us to solve a number of problems that previous proposals faced. Having two content norms has allowed us to vindicate Holmes' judgment of Watson's belief, thus explaining the disagreement. Importantly, it does this in a way that generalizes to all instances of disagreement, modal and otherwise, while simultaneously explaining the manner in which modal disagreements are like, and unlike more straightforward instances of disagreement. However, I have yet to explain how forced decision is avoided. According to my

proposal, undecided agents are informationally incorrect. Moreover, consciously undecided agents should recognize themselves to be informationally incorrect. What prevents undecided agents like Watson from guessing?

The answer arises from the fact that locative incorrectness is a special case of informational incorrectness. It follows that agents who believe falsely are locatively and informationally incorrect. Undecided agents are merely informationally incorrect. This is to say that agents who believe falsely aren't incorrect in an entirely different way than agents who are undecided. Both are informationally incorrect insofar as they are ignorant. However, agents who believe falsely also misrepresent the way the world is. Accordingly, agents who are locatively incorrect are incorrect in two ways, whereas agents who are informationally incorrect only suffer from one form of incorrectness.[39] This suggests that locatively incorrect agents are, in some sense, doxastically worse off than agents who are merely informationally incorrect. This captures the intuitive manner in which agents who believe falsely are in an inferior doxastic position than agents who remain undecided. This distinction is not only intuitive, but it is antecedently suggested by the test semantics. Agents like Lestrade who hold locatively incorrect beliefs must revise their information state *before* their state can be truthful. Without this, update with the facts will always yield a crash since for any s, if $s \vDash m$ then $s[\neg m] = \varnothing$. Watson, however, does not need to revise his information before update with $\neg m$ and the update itself will disabuse him of his informationally incorrect belief that Might m.

Thus, the test semantics, as well as the norms I've proposed, suggest a principle of *doxastic conservativity*.

Doxastic Conservativity: It is better to be merely informationally correct than to be locatively incorrect.

[39] One might instead say that locatively incorrect agents are wrong in the same way as merely informationally incorrect ones, only worse. Either conception works.

According to doxastic conservativity, it is better to be undecided than to believe falsely. This principle explains the aforementioned intuition that Lestrade's epistemic position is worse than Watson's. It also explains why undecided agents do not, and should not, guess when they realize that the content of their indecisive belief is informationally incorrect. Suppose a is consciously undecided about p and thus $I_a \vDash \text{Might } p \wedge \text{Might } \neg p$. a will recognize that the content of their belief is informationally incorrect, however, this will not compel them to guess. In first place, the homogeneity of error no longer holds, and not all incorrect contents are as wrong as false contents. Accordingly, not all incorrectness judgments motivate immediate belief revision. Secondly, two kinds of correctness ensure that there is now substantive risk associated with guessing. With strong correctness, a guess always increased the likelihood of correctness. Once locative and informational incorrectness are distinguished, this no longer holds. If a guesses right then they avoid incorrectness altogether, but if they guess wrong, then they will be locatively and informationally incorrect. Better, then, to remain undecided, and thus, forced decision is avoided.

The final problem to be addressed is generalized error attribution, according to which, all agents, including undecided agents, should uniformly judge indecisive contents to be incorrect. The principle still holds within the present framework, but it ceases to be problematic. According to informational correctness, agents will judge indecisive contents, including contents that they, themselves, believe, to be merely informationally incorrect. Mere informational correctness maps to failure to believe truly, which is precisely what agents who believe indecisive contents do. There is thus nothing wrong with recognition that indecision is incorrect, now that the homogeneity of error no longer holds.

This also dissolves the paradox in (3.2). For strong correctness, there was only one kind of wrong, and it was the same kind as believing falsely. Thus it struck as paradoxical to hold a belief while also judging it to be incorrect. According to the new

proposal, judgements of incorrectness are associated with informational incorrectness, and thus, it is perfectly coherent to hold an indecisive belief while also recognizing that one fails to believe truly. This nicely explains away the paradox, but it does not explain why utterances like (3.2) are infelicitous. It similarly fails to explain why replies like (37c) are infelicitous. This is to say that distinguishing between informational and locative correctness explains why an agent can hold an indecisive belief while judging it to be wrong, but it doesn't explain why undecided agents can't *say* that indecisive contents are wrong. Since this problem concerns assertion, a pragmatic principle is appropriate.

Informative Judgment: Only express judgments about content when they are *informative*.

According to informative judgment, cooperative conversational participants should only express normative judgments about content when they have nontrivial information to convey. We can illustrate this with Holmes, who can felicitously assert that the content of both Watson and Lestrade's respective beliefs are incorrect. Holmes does not merely express this judgment, however. He also asserts $\neg m$. Holmes has substantive information to convey, and his judgment comes bundled with it. Notice that it would be uncooperative/infelicitous for Holmes to merely assert that Watson was incorrect, without further comment.

Things are different for Watson, who is undecided, and has no information for Holmes. Thus, even if Watson does not accept Holmes' belief, and remains undecided, he will not express an incorrectness judgement about the contents of Holmes' belief.[40] Turning our attention to the puzzle at hand, while undecided agents will recognize indecisive contents to be informationally incorrect, they will never have anything informative to say, and so they tend to keep normative judgments, positive

[40]Any negative normative judgment Watson could express would concern formative correctness. So long has he is undecided, he cannot judge the content of Holmes' belief to be wrong.

or negative, to themselves. Thus, informative judgement explains the infelicity of (3.2) and (37c). This principle, coupled with locative and informational correctness, defangs the problems brought on by generalized error attribution. Moreover, the principle seems to be a specific instance of a more general pragmatic principle suggesting that cooperative interlocutors should keep their utterances informative.

Appeal to informational and locative correctness yields the requisite correctness judgements from interlocutors like Holmes. In addition, it avoids all of the problems associated with competing views. It also remains suitably objective, as both informational and locative correctness measure correctness against the facts that populate the actual world. It distinguishes between the incorrectness of doxastic indecision, and the incorrectness of believing falsely. In addition, the separation of these two kinds of correctness is highly plausible and manifests itself in conversations in ways that the laymen is sensitive to. Best of all, extant developments of the test semantics can adopt these correctness conditions for dynamic contents without making other substantive changes to their views. For instance, the lion's share of Willer (2013), including the entirety of the formal machinery, can remain intact. One need only replace evidential correctness with locative and informational correctness, and one can avoid the problems I've discussed. Things are even better for Lennertz (2019), which argues that there are cases of asymmetric disagreement. According to Lennertz, Holmes disagrees with Watson, but Watson does not disagree with Holmes. The arguments made presently are not merely consistent with Lennertz' position but appear to support it. Note that in virtue of being undecided, Watson does not know whether Holmes is locatively (and thereby informationally) correct or incorrect, since Watson does not know the facts. Meanwhile, Holmes thinks Watson is informationally incorrect and disagrees. If disagreements concerning content come bundled with judgments about correctness, then this seems to vindicate Lennertz' claim that Watson does not disagree with Holmes, despite the fact that Holmes disagrees with him. It offers

further support by bolstering the distinction between asymmetric disagreements like that between Holmes and Watson, and symmetric disagreements like that between Holmes and Lestrade. Holmes thinks Watson is merely informationally incorrect, while Holmes thinks Lestrade is locatively incorrect. Thus, I conclude that proponents of the test semantics should integrate locative and informational correctness into their proposed frameworks.

2.5 Concluding Remarks

A well-worn metaphor in epistemology analogizes belief and the firing of an arrow. Belief aims at truth in the way that the archer aims at his target. In traditional theories, indecision is the absence of belief, and the arrow is loosed only when a belief is formed. To believe truly is to strike the bullseye. The analogy, however, begins to unravel when applied to previous developments of the test semantics. According to previous accounts, propositional beliefs still aim at truth, but non-propositional beliefs fit poorly into the analogy, as their relationship with truth is indirect. Despite this, they are considered beliefs, proper, and are somehow held to the same standard of correctness as those that directly strive for truth. It is, by my lights, unclear how this fits into an otherwise attractive analogy.

Appeal to locative and informational correctness, in addition to generating the requisite correctness judgments on the part of Holmes, allows the analogy to apply clearly and succinctly to the test semantics. Locative contents indeed aim at the truth, and they hit the target when they are locatively correct. Non-locative beliefs do not fire an arrow, but instead merely take aim at the truth. They are informationally correct when the archer sets his sights on the target. Thus, belief still aims at truth, but not all beliefs loose an arrow.

Metaphors aside, my proposal is fairly simple. The test semantics allows for

two distinct kinds of contents. Things like indecision, which are classically treated as the absence of content, now have content, and so we should not be surprised when a single content norm fails to appropriately govern these two kinds of content. Things are further complexified by the fact that representational contents and non-representational contents have various, often asymmetric, semantic relations with one another. My proposal addresses these issues by insisting that content should be governed by content norms (as opposed to formative ones) and proposing two such norms: informational and locative correctness, and finally characterizing the interactions therebetween. Again, we should not be surprised that problems arise when we try to force formative norms to play a role that they simply cannot play. I take this to be the moral of the story, and I take this moral to be more significant than the details of locative and informational correctness. I also expect that this moral can be fruitfully applied to other non-propositional semantic accounts as well. Even if the details differ, we do well to admit of correctness conditions for non-truth-conditional contents that holds them to standards which suit the role they play in our thought and talk.

Chapter 3

Context, Contradiction, and Consistency

Synopsis:

Dynamic semantics violates numerous classical laws, including Non-Contradiction. Proponents of dynamic semantics have offered no explanation for this behavior, and some critics consider this result to be strong evidence against the tenability of the dynamic program. I defend and explain failures of Non-Contradiction by comparing dynamic semantics and classical, truth-conditional semantics in terms of their idealizing assumptions. I demonstrate that dynamic semantics rejects context fixity, an idealizing assumption that truth-conditional semantics typically adopts. I then argue that any semantics which rejects context fixity should, by the classical semanticists own lights, violate non-contradiction under certain circumstances. I then demonstrate that dynamic semantics violates Non-Contradiction in all and only those circumstances. I then consider further indirect evidence in support of this result. I close by suggesting that discussion of idealizing assumptions, common in the sciences, is similarly crucial to fruitful discussion in natural language semantics.

Introduction

Dynamic semantics for epistemic modals is non-classical in a variety of respects. In addition to violating all of the structural rules commonly associated with classical entailment, a laundry list of classically valid inference patterns are invalidated. Consider an instance of each below.

(1) **Commutativity**: $\varphi \wedge \psi \Rightarrow \psi \wedge \varphi$.

(2) **Monotonicity**: If $\Gamma \Rightarrow \varphi$ then $\Gamma \cup \Gamma' \Rightarrow \varphi$.

Commutativity and Monotonicity both hold in classical logic, but fail in dynamic semantics. These failures enjoy somewhat privileged status among the classical rules violated, since they are motivated by specific features of natural language. Commutativity fails due to putatively asymmetric behaviors of natural language conjunction. Failures of monotonicity capture the fact that utterances of epistemically modalized sentences are infelicitous in contexts where the negation of the prejacent is established.

Not all deviations from classical logic receive such direct explanations, however. In fact, most receive none at all.

(3) **Non-Contradiction**: $\varphi \wedge \neg\varphi$ is inconsistent.

Like Commutativity and Monotonicity, Non-Contradiction holds in classical logic, but does not hold in dynamic semantics. Unlike Commutativity and Monotonicity, violations of Non-Contradiction, while recognized in the literature, are not explicitly argued for, nor are any motivating cases from natural language proffered. It thus remains unclear whether, and if so, why, we should accept failures of Non-Contradiction in a semantics for natural language. One might be tempted to search for natural language cases that support the rejection of Non-Contradiction. In particular, some sentence of the form $\varphi \wedge \neg\varphi$ that is felicitously utterable would provide justification in similar style to those provided for violations of Commutativity and Monotonicity. Unfortunately, such cases are not likely forthcoming. The reason, in brief, is

that dynamic semantics only violates Non-Contradiction when substitution instances for φ are *not idempotent*. Among the simplest examples of formula that are not idempotent is $\Diamond p \wedge \neg p$, which means that among the simplest counterexamples to Non-Contradiction is:

(4) $(\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p)$

(4) has a degree of complexity that makes utterances of its natural language correlates clunky, extremely ambiguous, and unlike any commonly spoken sentences of natural language. This means that *even if* the dynamic semanticist were to successfully argue for the consistency of such sentences, it would be nearly impossible to provide an instance of a felicitous utterance thereof. This is to say that any potentially consistent utterances of sentences of the form $\varphi \wedge \neg\varphi$ are likely to be infelicitous for other reasons.[1] As such, direct and convincing linguistic motivation to reject Non-Contradiction is unlikely to be found.

Failures of Non-Contradiction may be surprising and unintuitive, and the lack of explanation may be considered a mark against the dynamic program. Matters are worsened by the fact that violations of Non-Contradiction are not alone, and that much of the non-classical behavior of dynamic semantics faces the same, or similar, explanatory problems.[2] As such, the proponent of dynamic semantics owes a substantial explanatory debt, and it is far from obvious what adequate payment would look like. The present paper sketches the contours of a satisfying explanation and aims to make a substantial down payment. To do this, I argue that failures of Non-

[1] This is exacerbated by the fact that (4) is, by the dynamic semanticist's own lights, incoherent. Incoherent sentences are not necessarily inconsistent (they don't crash every information state), but there is no state that supports them. Dynamic semanticists have often appealed to incoherence to explain various instances of non-catastrophic infelicity. As such, I would expect (4) to be at least marginally infelicitous. See §1.1.2 for the formal definition of coherence.

[2] For instance, Excluded Middle is also violated, when its substitution instances are not idempotent. As with violations of Non-Contradiction, intuitions on glosses that support rejection of the principle are incredibly difficult to come by. For simplicity, we restrict our attention to Non-Contradiction, and omit disjunction entirely. However, I expect many of the forthcoming considerations to apply to Excluded Middle as well.

Contradiction can be explained and justified once dynamic semantics is characterized in terms of its *idealizing assumptions*.

I begin by articulating an idealizing assumption called *context fixity*, which goes back, at least, to Kaplan and is operative in canonical accounts of context sensitivity. Kaplan observed that spatial and temporal shifts in context that occur mid-sentence would result in immediate failures of numerous classical laws. To avoid this, Kaplan famously argued that semantics concerns occurrences, rather than utterances, and proposed that we evaluate sentences and entire arguments at single contexts, that we might maintain classically valid inference patterns. I then demonstrate that dynamic semantics rejects context fixity, by allowing updates, rather than temporal or spatial shifts, to alter context. This allows for dynamic semantics to reject context fixity without resorting to theorizing in terms of utterances. I then consider some results concerning precisely when dynamic semantics predicts failures of Non-Contradiction and demonstrate that these failures always and only occur as a result of the kind of context shift Kaplan aimed to avoid. Thus, dynamic semantics makes predictions that are perfectly in line with the intuitions that motivate traditional contextualist approach to truth-conditional semantics in non-ideal environments. I conclude that dynamic semantics is justified in rejecting Non-Contradiction despite the fact that they are not motivated by any specific sentence glosses. I speculate that the same, or similar, explanations can plausibly be ported to other features of dynamic semantics that do not enjoy direct linguistic support. I also suggest that distinguishing semantic programs in terms of their idealizing assumptions allows us to understand putatively competing theories in considerably less antagonistic terms. This, in turn, allows us to more properly adjudicate differing predictions made by such frameworks.

3.1 Dynamic Semantics

3.1.1 Overview

The received view in natural language semantics takes the meaning of a sentence to be its truth-conditions, and semantic values are characterized in terms of propositions. Truth-conditional theories are mum with respect to the manner in which the meaning of a sentence uttered in context affects a state of information or conversation, as canonically, this is a job for pragmatics.[3] While the truth-conditional theorist has a litany of options for addressing this issue, adoption of truth-conditional semantics does not come with any particular commitments about how this process is carried out. Dynamic semantics for epistemic modals distinguishes itself from truth-conditional semantics by characterizing the compositional semantic values of sentences in terms of their discourse effects, rather than their truth-conditions. Instead, the meaning of a sentence is its capacity to change a state of information: a *context change potential* (CCP). Accordingly, the semantic values of the connectives are similarly defined in terms of CCPs. Such semantics can be called *compositionally dynamic*, which is to say that the discourse effects of sentences are determined by the compositional structure of the sentence itself, (Rothschild and Yalcin, 2016, pp.334). As each component of a sentence is processed, it updates the state of information to reflect this change. Since discourse effects are calculated compositionally, if a single sentence's compositional value is constituted by multiple sequential updates, the conversational context upon which later updates apply is not guaranteed to be the same as the initial context; context changes mid-sentence. This results in what can be called a *local context*, an intermediate state of information that reflects some, but not all, of the updates contained in the meaning of a given sentence, (Dever, 2013, pp.113). While a dynamic semantics is not required in order to appeal to local contexts, dynamic semantics

[3]See Stalnaker (1978) for the canonical account.

are unique insofar as local contexts determine the compositional semantic values of sentences.[4]

3.1.2 Formal Details

I now consider the propositional fragment of the dynamic language proposed in Groenendijk and Stokhof (1991b).[5]

Let $\mathcal{L}_1$ be generated by the grammar:

$$\varphi ::= \mathrm{p} \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi$$

Dynamic semantics is intended to capture the manner in which sentences can update a state of information based on their compositional structure. We thus characterize a general framework for the way in which sentences can update states of information. This can be captured by an *update system*.

(5) **Update System**: $\langle \mathcal{L}, C, \cdot[\cdot] \rangle$ is an update system if and only if $\mathcal{L}$ is a set of sentences, C is a set of information states, and $\cdot[\cdot]$ is a function which maps sentences of $\mathcal{L}$ to operations on C.

An update system models the manner in which sentences of a given language can change a state of information. Let W be the set of all functions $\mathcal{A} \mapsto \{0, 1\}$. Our update system will be $\langle \mathcal{L}_1, \mathcal{P}(W), \cdot[\cdot] \rangle$. Some information state $s \in \mathcal{P}(W)$ is a set of possible worlds compatible with the information contained within the state. We characterize our semantics for $\mathcal{L}_1$ in terms of operations on $\mathcal{P}(W)$.[6]

[4] See Schlenker (2009) for a pragmatic account of local contexts.

[5] See Veltman (1996), Groenendijk and Stokhof (1991a) van der Does et al. (1997), Aloni (2001), Gillies (2004), von Fintel and Gillies (2007), Yalcin (2012b), Yalcin (2015), Willer (2013), and Lennertz (2019) among many others, for developments and variations of the test semantics.

[6] According to Rothschild and Yalcin, a conversation system coupled with a semantics for the language determine what is called a *state system*. Whether, and in what capacity, a semantics is dynamic is determined by features of the state system. On this characterization, our present semantics is dynamic precisely because it fails to be idempotent and commutative. More on this presently.

(6) **Semantic Clauses**[7]

1. $s[\text{p}] = \{w \in s | w(\text{p}) = 1\}$
2. $s[\varphi \wedge \psi] = s[\varphi][\psi]$
3. $s[\neg\varphi] = s - s[\varphi]$
4. $s[\Diamond\varphi] = \{w \in s : s[\varphi] \neq \varnothing\}$

Updating a state with an atomic formula takes the intersection of the worlds in the state, and the set of worlds where that atomic is true. Conjunction is consecutive update, and negation takes the difference between the initial state, and what would result by updating the initial state with the non-negated formula.[8] Epistemic possibility (epistemic "might") performs a test, where the initial state is returned if the state is compatible with the prejacent of the modal. Otherwise, the test is "failed" and the absurd state $\varnothing$ is returned.

We can next define what it is for a state to support some information.

(7) **Support**: $s \vDash \varphi \Leftrightarrow s[\varphi] = s$

A state supports some information if and only if updating the state with that information does not change the state. Next, we define consequence and consistency.

(8) **Dynamic Consequence**: $\psi_1, ..., \psi_n \Vdash \varphi \Leftrightarrow \forall s, s[\psi_1]...[\psi_n] \vDash \varphi$

(9) **Dynamic Consistency**: $\psi_1, ..., \psi_n$ is consistent $\Leftrightarrow \exists s, s[\psi_1]...[\psi_n] \neq \varnothing$

(10) **Dynamic Inconsistency**: $\psi_1, ..., \psi_n$ is inconsistent $\Leftrightarrow \psi_1, ..., \psi_n$ is not consistent

[7] The present formulation only considers negation, conjunction, and epistemic possibility operators. Disjunction and quantification are suppressed, as the problem can be formulated in their absence.

[8] The failure of Non-Contradiction can be avoided by adopting the "static" versions of conjunction, which is set intersection. An example of this strategy can be found in the early pages of Veltman (1996). The problem, as Mandelkern (2020) observes, is that the static approach also prevents dynamic semantics from predicting many of the phenomena it is intended to predict, as the set theoretic operations are commutative.

A sequence of formula dynamically entails some formula φ if and only if every state updated with that sequence will support φ. Similarly, a sequence of updates is dynamically consistent if and only if there is a state that can be updated with that sequence without resulting in the absurd state. Note that this notion of entailment does not concern truth, and instead captures support preservation. Also note that the above notions are order sensitive. For instance, $\langle \Diamond p, \neg p \rangle$ is dynamically consistent while $\langle \neg p, \Diamond p \rangle$ is not.

While I've elected to define dynamic entailment and consistency as above, non-equivalent alternatives are both available and useful. A variant of inconsistency that is interdefinable with a different notion of entailment is also often appealed to.

(11) **Coherence**: $\psi_1, ..., \psi_n$ is coherent $\Leftrightarrow \exists s \neq \varnothing$, s.t. $s \vDash \psi_1, ..., s \vDash \psi_n$

Unlike consistency, coherence demands that a single state support each update individually, rather than in sequence. Unlike consistency, coherence is not order sensitive, e.g. neither $\langle \Diamond p, \neg p \rangle$ nor $\langle \neg p, \Diamond p \rangle$ are coherent. Any update that is not coherent (incoherent) is not supported by any single state.

A few more definitions will be useful shortly.

(12) **Idempotence**: $[\varphi]$ is idempotent $\Leftrightarrow \forall s$, $s[\varphi] = s[\varphi][\varphi]$

An update is idempotent if and only if subsequent applications of the same update do not yield any change beyond the first. Importantly, a certain class of updates in this semantics are not idempotent.[9]

(13) **Distributivity**: $[\varphi]$ is distributive $\Leftrightarrow$ for all $s, s[\varphi] = \bigcup\{\{w\}[\varphi] | w \in s\}$

An update is distributive just in case updating a state with the formula is equivalent to the union of the individual updates of the singleton set of each world in that state, (van Benthem, 1989, pp.364). All contents that do not include the $\Diamond$ operator are

[9]Properly speaking, updates, rather than formula, have the property of idempotence. For brevity, however, we may sometimes say that "φ is idempotent" to mean that "$[\varphi]$ is idempotent."

distributive. Distributive update can always be characterized by the union of an information state, and a fixed set of worlds. This set of worlds will not change depending on the nature of the state it updates, and thus, updates that are distributive can be said to express propositions. Even though the CCP associated with the update is not, itself, a proposition, a proposition can always be recovered from the update. Like failures of idempotence, failures of distributivity are associated with semantics that are dynamic. In addition, all updates that are distributive are idempotent. Some updates that fail to be distributive are not idempotent. Alternatively, updates that are not distributive cannot be associated with a particular proposition; updating with such formula will not always intersect with a fixed set of worlds, since the update is sensitive to the state it updates.

3.1.3 Selected Applications

The above semantics has been used to explain various behaviors of natural language, often involving epistemic modals and conjunction, that putatively fail to conform to patterns of classical logic. Here, we rehearse a select few that will be relevant for discussion.[10] Veltman (1996) observes a discrepancy in the following:

(14) ?? Adam might be at the door and...Adam isn't at the door.

(15) # Adam isn't at the door, and he might be.

(14) is marginally assertible at certain contexts, namely, those where one expects a particular visitor but is met with another. (15), however, is brutely infelicitous across all contexts. (14) is of the form $\Diamond p \wedge \neg p$ while (15) is of the form $\neg p \wedge \Diamond p$. This discrepancy can be explained by the fact that (14) is consistent, but (15) is

[10]Dynamic semantics has been commonly used to solve modal disagreements, as well as predicting the infelicity of epistemic contradictions. See Willer (2013) for a dynamic solution to both. See MacFarlane (2011) and MacFarlane (2014) for discussions of modal disagreement and see Yalcin (2007) for discussion of epistemic contradictions.

inconsistent.[11] Of course, the only difference between (14) and (15) is the order of the conjuncts. Dynamic semantics can explain this discrepancy due, in part, to the failure of commutativity of conjunction. Cases such as these also motivate failures of monotonicity. For instance, an agent undecided about whether p can assert or believe $\Diamond p$. However, should an agent learn that p is false and come to believe $\neg p$ they can no longer maintain $\Diamond p$. This suggests, contra the classical rule of monotonicity, that gaining certain information results in the loss of other information. This also helps explain the discrepancy in infelicity observed between (14) and (15).

While (14) is marginally assertible, it is observed that (14) is never felicitously utterable twice in succession:

(16) # Adam might be at the door, and he isn't at the door, and he might be, and he isn't.

(16) is of the form $(\Diamond p \wedge \neg p) \wedge (\Diamond p \wedge \neg p)$, so according to dynamic semantics, its meaning is captured by the following sequence of updates: $[\Diamond p][\neg p][\Diamond p][\neg p]$. No state, however, can sustain sequential update with the middle two updates, $[\neg p][\Diamond p]$, without resulting in the absurd state. Thus, (16) is inconsistent and thereby infelicitous. In addition to the infelicity of (16), dynamic semantics also explains why the first instance of (14) can have a distinct discourse effect from a second. $[\Diamond p \wedge \neg p]$ is not idempotent, meaning that subsequent updates can have effects over and above the first. Thus, cases like (16) vindicate the presence of non-idempotent update in dynamic semantics. So-called, static, frameworks are unable to predict the infelicity of such cases semantically and will have to appeal to pragmatics.

Dynamic semantics also predicts related behaviors including modal disagreements

[11] This discrepancy in infelicity is also captured in data gathered Knobe and Yalcin (2014). In addition, dynamic semantics can predict the marginal infelicity of (14) in virtue of the fact that it is incoherent, despite the fact that it is consistent. This is to say that a state can sustain update with (14) without crashing, but no single state can support each update in (14). This explains why (14) sounds funny, but isn't always catastrophically infelicitous.

and epistemic contradictions.[12] Further, the dynamic clause for conjunction allows for the semantics to be implemented in other frameworks designed to capture other dynamic phenomena, including presupposition projection and anaphora resolution. The important point, however, is that the motivations for the rejection of Commutativity and Monotonicity are clear, explicit, and motivated by concrete examples involving belief or assertion. The literature, however, does not contain any motivating examples for violations of Non-Contradiction.

3.2 Violations of Non-Contradiction

Dynamic violates Non-Contradiction, as well as several other classical laws. A violation of Non-Contradiction means that there is some state, s and some formula φ such that $s[\varphi \wedge \neg\varphi] \neq \varnothing$. As demonstrated in Mandelkern (2020), Non-Contradiction fails when $[\varphi]$ is not idempotent. One of the simplest examples of non-idempotence is:

(17) $[\Diamond p \wedge \neg p]$[13]

If we use $\Diamond p \wedge \neg p$ as a substitution instance for $\varphi \wedge \neg\varphi$, we get:

(18) $(\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p)$

Despite being of the form $\varphi \wedge \neg\varphi$, (18) is consistent, since there exists a state that can sustain update with (18) without resulting in the absurd state.[14] Therefore, it is not the case that all substitution instances of $\varphi \wedge \neg\varphi$ are contradictions and Non-Contradiction does not hold.

Importantly, not all instances of $\varphi \wedge \neg\varphi$ are consistent, and it fails in a very particular set of circumstances. While Mandelkern (2020) proves that $[\varphi]$ is not

[12] See Willer (2013), Willer (2015), and Lennertz (2019) for examples.

[13] To show that $\Diamond p \wedge \neg p$ is not idempotent, let s be an information state s.t. $s = \{w, w'\}$, $w(p) = 1$, and $w'(p) = 0$. From the compositional values of the connectives, we can see that $s[\Diamond p \wedge \neg p] = \{w'\}$, while $s[\Diamond p \wedge \neg p][\Diamond p \wedge \neg p] = \varnothing$. Therefore, $s[\Diamond p \wedge \neg p] \neq s[\Diamond p \wedge \neg p][\Diamond p \wedge \neg p]$ and $[\Diamond p \wedge \neg p]$ is not idempotent.

[14] Proof in Appendix i.

idempotent only if $\varphi \wedge \neg\varphi$ is consistent, the converse also holds.[15] Thus, $\varphi \wedge \neg\varphi$ is consistent if and only if $[\varphi]$ is not idempotent. This is to say that failures of Non-Contradiction not only include substitution instances where $[\varphi]$ is idempotent, but are exhausted by them. This means, *inter alia*, that for idempotent updates, Non-Contradiction holds exactly as it does in the classical case. This includes, but is not limited to, all distributive (propositional) formula, meaning that for any $[\varphi]$ that is distributive, $\varphi \wedge \neg\varphi$ is inconsistent. Thus, for any substitution instances that are truth-evaluable, including atomics, Non-Contradiction still holds. $p \wedge \neg p$, for example, is inconsistent. This explains why many instances of $\varphi \wedge \neg\varphi$ are straightforwardly infelicitous.

(19) # San Francisco is in California and San Francisco is not in California.

Thus, dynamic semantics does not lose its capacity to predict that direct contradictions of truth-evaluable-contents are infelicitous. In addition, this also ensures that dynamic semantics does not recognize anything resembling a true contradiction. While there are instances of $\varphi \wedge \neg\varphi$ that are consistent, no instance that is consistent will express a proposition. In this case, recognizing the failure of Non-Contradiction does not commit one to true contradictions, but the substantially more conservative claim certain states of information can be updated by certain formula of the form $\varphi \wedge \neg\varphi$ with crashing.

Formula of the form $\varphi \wedge \neg\varphi$ that are consistent must contain instances of φ s.t. $[\varphi]$ is not idempotent. Crucially, an update is not idempotent only if it is not distributive, and all non-distributive updates contain at least one instance of the epistemic possibility operator $\Diamond$, thus, all non-idempotent updates will as well. In addition, for any formula φ, if φ is of the form $\Diamond\psi$, then $[\varphi]$ is idempotent.[16] As such, any formula that is not idempotent must contain $\Diamond$ embedded under another operator. It will not

[15] Proof in Appendix i.
[16] Proof in Appendix i.

be enough to embedded the modal under negation, since for all φ, $[\neg\Diamond\varphi]$ is idempotent.[17] Moreover, since double negation elimination remains valid, no series of iterated negations will do the trick. Instead, the modal will need to be embedded under conjunction.[18] Lastly, since any sequence of tests and conjunctions without negation will be idempotent, the formula must also contain a negation. The result is that any non-idempotent formula must minimally contain at least on instance of all three connectives. This, in turn, means that (17) above is the simplest non-idempotent update we can expect to find. Since Non-Contradiction involves two instances of φ embedded under conjunction and negation, this means that the simplest failures of Non-Contradiction will include no fewer than eight operators, which means that within the test semantics, (18) is the simplest failure of Non-Contradiction available. It's natural language correlate is as follows.

(20) ?? (Adam might be in his office, and Adam is not in his office) and it is not the case that (Adam might be in his office, and Adam is not in his office.)

The above is, literally and figuratively, difficult to parse, and it would be nearly impossible without the help of the added parentheses. Moreover, alternative failures of Non-Contradiction will have, at minimum, the same level of complexity. Utterances thereof would seem, at the very least, to flout the maxim of manner. Nonetheless, according to dynamic semantics, such sentences are consistent. While it does not follow that dynamic semanticists are committed to the claim that such sentences are felicitously utterable, it does mean that dynamic semantics cannot appeal to their inconsistency to predict any infelicity, if such there be.

As we saw in §3.1.3, failures of Commutativity and Monotonicity were motivated by intuitions concerning reasonably straightforward utterances and inferences. To

[17] Proof in Appendix i.

[18] Non-idempotent updates can also be generated with disjunction, but I set these aside since we do not currently consider disjunction. More generally, however, non-idempotent updates will require at least one binary operator.

defend failures of Non-Contradiction in similar fashion would require something like a marginally assertable instance of (20). However, given the sentence's complexity, as well as the fact that on the required disambiguation, the sentence is not coherent, it is profoundly unlikely that such an example can be provided. Even if such sentences should be treated as semantically consistent, they will often be infelicitous or confusing for other reasons, and intuitions will be muddled, at best. There is not, to my knowledge, any examples or intuitions involving speaker data that motivate this result in the literature. Given the complexity of the failure cases, it will be difficult, if not impossible, to motivate failures of Non-Contradiction with intuitive judgments about assertions in context. This means that failures of Non-Contradiction not only remain unexplained, but that the standard dynamic strategy of motivating non-classical semantics with examples of utterances and discourses in context is of no help here. The dynamic semanticist will need to explain this result by some other means.

3.3 Context and Idealization

To understand what to make of this, it will be instructive to consider the different ways in which truth-conditional and dynamic semantic theories interact with context, as well as the insights that motivate these choices.

3.3.1 Context, Canonically

The standard approach regarding semantic interactions with context is Kaplanian in spirit, if not in letter. Kaplan, motivated primarily by demonstratives and indexicals, argues that the denotation of a sentence is relativized to a context and an index. For Kaplan, a context is a four tuple of an agent, time, location, and possible world. An index is an n-tuple with arbitrarily many parameters that model any other information that a sentence might be sensitive to. It is standardly taken, however,

that the information in a Kaplanian context should be enough to fully determine any information for an index, so the index is often suppressed.[19] Thus, it is pedestrian to relativize the denotation of sentence to a context c which supplies all of the relevant contextually supplied information, like so.

(21) $[\![\varphi]\!]_c$

According to this approach, sentences are evaluated at single contexts, and any components of φ with contextually supplied values will have those values supplied by c. Kaplan further posits the notion of *truth at a context* and defines validity as preservation thereof. Thus, not merely sentences, but entire arguments are evaluated relative to a single, fixed, context. This strategy can be captured by the following principle.

(22) **Context Fixity**: All components of a sentence (or argument) are evaluated relative to the same context.

According to Context Fixity, contexts do not shift mid-sentence. Each subformula of a sentence that requires any contextually supplied values will receive these values from the same context. The same goes for different premises in an argument, all of which are evaluated at a single context.[20] It is easy, however, to imagine cases of language use where context does change mid-sentence, or mid-argument. Kaplan recognizes this possibility, and is careful to distinguish between an *occurrence* and an *utterance*. An occurrence of a sentence at a context is simply a sentence context pair. We can evaluate an occurrence of a sentence at a context, and determine whether it is true, regardless of whether the sentence is uttered in that context (or whether the language of the sentence even exists at that context, etc.). Utterances, on the other

[19] Since a context should fully determine an index, Lewis (1980) famously calls Kaplan's distinction between context and index a "distinction without a difference."

[20] The strategy of defining entailment in terms of truth at a context is ubiquitous in post-Kaplan approaches to context sensitivity as well. See Predelli (2005) for an example. Predelli, contra Lewis, maintains Kaplan's distinction between context and index, and thus treats entailment as preservation of truth at a context and an index. This is relevant because, while we use the Lewisian strategy for simplicity, nothing hangs on it.

hand, are speech acts, about which Kaplan says the following:

> Utterances take time, and are produced one at a time; this will not do for an analysis of validity. By the time an agent finished uttering a very, very long true premise, the premise may have gone false. Thus even the most trivial inferences, P therefore P, may appear invalid. Also, there are sentences which express a truth in certain contexts, but not if uttered. For example "I say nothing." Logic and semantics are concerned not with the vagaries of actions, but the verities of meanings, (Kaplan, 1989, pp.584).

Kaplan believes that occurrences, rather than utterances, are the proper subject matter of semantics. Even though no two spoken sentences can be uttered in the exact same context, occurrences of these sentences can be evaluated at single, fixed contexts, and we can model whether they preserve truth at these contexts. In so doing, we can theorize entirely in terms of occurrences of sentences evaluated at the same context, and thus, validity that concerns preservation of truth at a context is an account of occurrence validity. According to Kaplan, if we do not do this, and theorize in such a way that context can change mid-sentence, we intuit immediate failures of inference patterns like repetition. To avoid dealing with situations like these, Kaplan thinks that we ought to characterize validity at the occurrence level, rather than the utterance level. Kaplan is extremely pessimistic about the prospects of the latter.

> I [Kaplan] am unclear even as to what arguments *ought* to come out as utterance-valid (as opposed to occurrence-valid). There are different notions of utterance-validity corresponding to different assumptions and idealizations. With no idealizations, the rules of repetition and double negation become invalid. This seems hopeless, (Kaplan, 1989, pp.585)

It is important to appreciate that Kaplan is not arguing that we adopt context fixity because it is true of natural language consequence. Indeed, it is almost certainly false, as any account of entailment which satisfies context fixity will fail to capture even the simplest cases of intercontextual entailment, e.g. If "It will rain tomorrow" is true at a context today, then "It will rain today." must be true at a context tomorrow.[21] This is not an objection to Kaplan's claims, but merely the observation that Kaplan's proposal is based on the intractability of the alternative. This is to say that Kaplan is proposing that we make an *idealizing assumption*. A characteristic feature of such assumptions is that they are not made based on their truth, real or believed. Instead, they are made for practical reasons like simplicity, understandability, and tractability, (Potochnik, 2017, pp.43-44). Famous examples of such assumptions include frictionless planes and completely rational agents. Despite the fact that neither exist, each has proven indispensable in our theorizing. The same may plausibly be said of context fixity, as no two sentences are ever uttered at a single context. It has been, nonetheless, extremely profitable and informative to use this idealization in understanding indexicals and demonstratives. It has also provided the possibility of a generalized framework for theorizing about context sensitive meanings. Thus, according to the Kaplanian position, we needn't deny that contexts can change mid-sentence. Rather, given the intractability of theorizing in terms of utterance validity, we can idealize by holding context fixed. We can then cleanly discuss entailments of occurrences.

As far as Kaplan is concerned, Context Fixity is an idealizing assumption that is foisted upon given the "hopelessness" of theorizing in its absence. Should we reject it, Kaplan believes that we must suffer the following:

1. We are forced to theorize in terms of utterances, rather than occurrences.

[21] Such examples are famously discussed in Frege (1918). See Zardini (2014) for a logic intended to capture these entailments. Not surprisingly, the system presented in Zardini (2014) rejects context fixity.

2. We have no clear intuitions about utterance validity.

3. We would lose patterns like Repetition.

Fortunately, things are not hopeless. 1. in particular, is false, and there is room in the logical space for a theory which characterizes validity at the level of occurrences, but rejects context fixity. We might choose to evaluate particular occurrences at a context, but allow subsequent sentences to modify the context. We may then characterize a notion of validity which accounts for information preservation through changing contexts. More on this in a moment, but the crucial point is that rejecting context fixity does not require that we theorize in terms of utterances. This also immediately absolves us of any responsibility to address 2. as we needn't discuss utterance validity at all. 3. remains, and patterns like Repetition would be lost, but this is far from hopeless. Rather, this is exactly what we would expect from a notion of entailment which rejects context fixity. Somewhat roughly, if different instances of the same sentence can express different meanings/contents/truth-conditions at different contexts, and we intend to capture a notion of validity that recognizes information preservation through context change, then we should expect failures of rules like Repetition. Put slightly differently, we would expect this account of entailment to reject the structural rule of *Reflexivity*.

(23) **Reflexivity**: $\varphi \Rightarrow \varphi$

As meaning and contexts change, we may lose information that we once had. A satisfactory logic of context change would need to predict and explain precisely when this would happen.

Thus, rejecting context fixity is far from hopeless, and does not require that we theorize in terms of utterances. We will, however, see some non-classical behaviors including failures of Reflexivity.[22] Importantly, by reasoning very similar to

[22]I mention that not all rejections of context fixity result in failures of reflexivity. An example

Kaplan's, we should also expect other non-classical behaviors, including failures of Non-Contradicton. Consider natural language sentences of the following form:

(24) φ_1 and it's not the case that φ_2.

Each instance of the sentence φ is given a metalinguistic marker in (24) above. It is easy to see that if φ_1 and φ_2 are evaluated at the same context, we understand (24) to be inconsistent. If φ_1 and φ_2 are evaluated at the same context, then they will 'mean the same thing' so to speak. Our intuitive understanding of negation and conjunction also tell us that something in conjunction with the negation of that very thing is contradictory, and thereby inconsistent. However, once we reject context fixity, we may evaluate φ_1 at a different context from φ_2, meaning that, at some level, φ_1 and φ_2 aren't guaranteed to have the same meaning. In this case, the very same intuitions about negation and conjunction yield a different conclusion, namely, that if the context has changed enough, (24) is consistent. This reasoning is not importantly different from Kaplan's claims concerning the failure of repetition. We may thus conclude that rejecting context fixity opens the door for failures of Non-Contradiction in the same way it did for failures of Reflexivity. Importantly, this conclusion seems to arise from the very same intuitions that motivate the canonical approach. Thus, by Kaplan's lights, as well as our own, we can expect failures of Reflexivity and Non-Contradiction in accounts that reject context fixity.

3.3.2 Dynamic Contexts

The previous section established that there is room for theorizing which recognizes context change while still focusing exclusively on occurrences. I have also argued that, by Kaplanian lights, we should expect such a system to violate Reflexivity and

would include the propositional fragment of the $\mathcal{L}_1$ discussed above. Rather, we expect failures of reflexivity when we allow for what we might call *non-trivial context change*. Context change is non-trivial just in case the difference in contexts is substantial enough such that the same formula can have distinct meanings at each context. When I say a semantics rejects context fixity, I properly mean that it allows for non-trivial context change.

Non-Contradiction. The punchline, of course, is that this region of the theoretical space is not uncharted wilderness and has been inhabited by dynamic semantics for quite some time. Dynamic semantics processes updates relative to a context (an information state) and outputs a new, updated, context. Subsequent updates, including updates in the same sentence or argument, are updated relative to the output of the previous update, which will often be different than the initial context. We called such intermediate contexts local contexts, and we mentioned that it is often taken that the semantic significance of local contexts is characteristic of dynamic semantics. Dynamic semantics still trades in occurrences, and an occurrence of a sentence at a context will update that context. However, the context at which a sentence occurs can be changed by update with the sentence itself, and not just by pragmatic factors. Dynamically speaking, context change is not merely an artifact of the temporal extension of speech acts. Within a dynamic framework, the meanings of occurrences themselves can change at least one contextual parameter: the information state. It should thus be clear that dynamic semantics directly rejects context fixity while still trading in terms of occurrences. Viewed from the Kaplanian perspective, the dynamic semantics we presently consider recognizes a privileged contextual parameter, the information state, and allows that parameter to shift as it is updated. Occurrences of sentences are still evaluated at a single input context, but the output of that update (and the input of subsequent updates) can be different. Every other contextual parameter may well still be handled in Kaplanian fashion. This difference, however, is enough to reject context fixity.

Since dynamic semantics rejects context fixity, and we should expect semantics which reject context fixity to violate Non-Contradiction and Reflexivity, we should expect dynamic semantics to violate both of these laws. This is exactly what we observe as neither Reflexivity nor Non-Contradiction hold in dynamic semantics, (van Benthem, 1996, 142). It may be obvious that a semantics which trades in context

change will reject context fixity. What may be less obvious is that when context fixity is viewed as an idealizing assumption, rather than an inviolable pillar of theorizing, it vastly changes what rejecting the principle amounts to. Dynamic semantics is often seen by supporters and critics alike as exotic, esoteric, and thoroughly punk rock, given its deviance from classical logic. It may be tempting to view the program this way when its characteristic feature, the rejection of context fixity, is a pillar of the canonical position. As much as I might enjoy this characterization, it is important to appreciate that the theoretical choice to reject context fixity seems far more temperate and benign once context fixity is recognized as not only an idealizing assumption, but a false one. This should not only change the way that we view and market dynamic semantics; it should also inform our understanding of the non-classical nature of dynamic entailment, at least broadly speaking. By Kaplan's own lights, holding different idealizing assumptions will yield different notions of validity. Accordingly, recognizing context fixity as an idealizing assumption not only makes the dynamic approach to entailment seem less fringe and extreme, but it also makes the classical approach significantly less sacrosanct.

3.4 Context and Contradiction

I have argued that we should expect failures of Non-Contradiction and Reflexivity in a dynamic framework. This does not, however, vindicate the specific failures of Non-Contradiction that dynamic semantics predicts. There is a profound disanalogy between between the spatial and temporal changes in context which concerned Kaplan, and the changes in information states operative in dynamic semantics. Dynamic semantics concerns occurrences of sentences and thus, sentences will be evaluated (as updates) at discrete contexts. This means that there will be none of the spatial and temporal shifting that made utterance validity so difficult to handle. Instead, within

dynamic semantics, the only way to change an information state is to update it with a formula. This places a significant constraint on when we can expect failures of Non-Contradiction. Consider formulas of $\mathcal{L}_1$ of the following form:

(25) $\varphi_1 \wedge \neg\varphi_2$

The kinds of failures of Non-Contradiction that concerned Kaplan would involve some temporal or spatial change that occurred between when φ_1 was uttered and the time that φ_2 was uttered. Kaplan's intuitions, however, can be generalized, and needn't involve time or space per se, and we should expect failures of (25) in cases where context, in whatever capacity, changes enough between update with φ_1 and update with φ_2 such that the meaning of φ_1 at its context does not entail the meaning of φ_2 at its context. Spatial and temporal changes to context are not recognized by our semantics, and thus, the kind of change that must be brought about between φ_1 and φ_2 that would result in a violation of Non-Contradiction can only come about by update. The clause for conjunction tells us that to update a state s with (25) is to update s with $[\varphi_1][\neg\varphi_2]$. What this means is that the only way for the initial context s to change before update with $\neg\varphi_2$ is by update with φ_1. Thus, without the help of changes in time and space, $[\varphi_1]$, itself, must change s enough so that $s[\varphi_1]$ can sustain update with $[\neg\varphi_2]$ without crashing. Any state s.t. $s \vDash \varphi$ cannot be updated with $\neg\varphi$ without crashing. Thus $[\varphi_1]$ must change the state enough s.t. $s[\varphi_1] \nvDash \varphi$. Slightly less formally, update with φ has to change the state enough that the output state does not support φ. This means that we should only expect failures of Non-Contradiction when substitution instances of φ fail to satisfy Reflexivity, i.e., $\varphi \nVdash \varphi$. This is to say that we would only expect (25) to be consistent in cases where φ does not entail itself.

Based on the intuitions about context change above, we should only expect the consistency of $\varphi \wedge \neg\varphi$ when φ violates reflexivity. It was also established, in §3.2, that $\varphi \wedge \neg\varphi$ is consistent if and only if $[\varphi]$ is not idempotent. It is easy to prove,

however, that these cases exactly coincide since $\varphi \not\Vdash \varphi$ if and only if $[\varphi]$ is not idempotent.[23] Thus, the same intuitions about context change which motivate the canonical view, vindicate dynamic semantics' predictions about failures of Non-Contradiction. Namely, we should expect the consistency of $\varphi \wedge \neg\varphi$ precisely when φ does not entail itself. Such cases, fairly intuitively, coincide with cases where repeated update with φ can bring about context changes over and above those brought about by the first. This explains why we should expect failures of Non-Contradiction to be directly associated with non-idempotent update. Such failures match, not at all coincidentally, with the failures predicted by dynamic semantics.

I demonstrated in §3.1.3 that the formal properties of idempotence and irreflexivity associated with $\Diamond p \wedge \neg p$ transfer to its natural language counterparts. (14) cannot be felicitously uttered twice in succession (non-idempotence) and moreover, anyone who honestly asserts (14) will not believe that Adam both might be at the door and isn't. Rather they will believe that Adam isn't at the door (irreflexivity).[24] By Kaplanian reasoning coupled with dynamic semantics, we thus expect sentences of the form φ and it's not the case that φ, when φ is substituted with (14), to be consistent. We expect Non-Contradiction to fail in all and only these cases, and this is exactly what dynamic semantics predicts.

The intuitions that motivate dynamic semantics, as well as the formal predictions it makes, dovetail perfectly with the Kaplanian intuitions about context change once we are willing to reject context fixity. The difference between the dynamic and the Kaplanian approach is that in dynamic frameworks, meanings themselves, and not only non-linguistic forces, can alter the contexts in which subsequent formula are evaluated. This allows for a separate avenue, unappreciated by Kaplan, for contexts to be changed by semantic forces alone. Kaplan's claims about meanings through changing contexts, however, remain pertinent. This is to say that the very failures of

[23] Proof in Appendix i.

[24] $\Diamond p \wedge \neg p \Vdash \neg p$ but $\Diamond p \wedge \neg p \not\Vdash \Diamond p \wedge \neg p$.

classical laws that Kaplan feared, are the very same violations of classical laws that dynamic semantics predicts and embraces.

3.5 What it Might Mean and What it Doesn't Mean

The previous sections have attempted to demonstrate that failures of Non-Contradiction follow from the dynamic rejection of context fixity coupled with broadly Kaplanian considerations about context sensitive meaning. The result is that formula of the form $\varphi \wedge \neg\varphi$ are consistent precisely when $[\varphi]$ is not idempotent. Thus, allies and critics alike should neither be confused nor surprised by this result. What to make of it is slightly less clear. Dynamic semantics is motivated by discourse patterns that are difficult to explain within canonical frameworks. The dynamic strategy is to treat discourses, rather than sentences, as the objects of semantic study. The result is a semantics with a notion of entailment that captures the manner in which information is preserved through changing contexts. While certain non-classical behaviors, like the infelicity of (15), are deliberately mirrored by certain formal features of the semantics, like the inconsistency of $\neg p \wedge \Diamond p$, we should not expect anything resembling a direct homomorphism between the two. This is to say that the fact that (18) is consistent does not mean that dynamic semantics predicts (20) to be felicitous. Rather, my argument merely absolves dynamic semantics of any obligation to treat it as consistent. Indeed, dynamic semantics seems committed to the claim that (20) is, at best, only marginally felicitous, given that it is incoherent. The point, however, is that since the dynamic semanticist rejects context fixity, they reject the claim that any formula of the form $\varphi \wedge \neg\varphi$ is bound to be inconsistent. Claims of infelicity based on intuitions that the dynamic semanticist explicitly rejects will remain unconvincing.

This also opens the door for a line of response to a related objection. Mandelkern

(2020) observes that while dynamic semantics treats (20) as infelicitous since (18) is incoherent, this sentence can be modalized to yield a sentence that is both consistent and coherent.

(26) $\Diamond((\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p))$

(27) ?? It might be the case that [(Adam might be in his office, and Adam is not in his office) and it is not the case that (Adam might be in his office, and Adam is not in his office.)]

The objection holds that, since the prejacent of (27) is clearly a contradiction, it should clearly be infelicitous. While this reasoning may be convincing to a proponent of classical semantics, since the dynamic semanticist rejects context fixity, the idea that $\varphi \wedge \neg\varphi$ is always a contradiction simply does not hold. Therefore, in certain cases, the update may well be acceptable. In the absence of proper data, and with the idealizing assumptions made explicit, this strikes as a conflict of intuitions, rather than an objection, proper.

I doubt such arguments will convince classical semanticists to jump ship. I do, however, expect these arguments to explain why dynamic semanticists remain similarly unmoved in light of classical criticism that dynamic semantics "fails" to validate Non-Contradiction. Brute insistence that dynamic semantics is obligated to treat $\varphi \wedge \neg\varphi$ as inconsistent either assumes a homomorphic relationship between inconsistency and infelicity, which is implausible, or assumes context fixity, which the dynamic semanticist directly rejects. In either case, the objection falls on deaf ears.

While my discussion has concerned Non-Contradiction, I expect the lesson to generalize to at least some of the other non-classical behavior that dynamic semantics exhibits. It is standard practice for theorists to recognize direct, particular non-classical behaviors of epistemic modals, and attempt to provide a semantics which is non-classical in only these respects. This is to say that the standard practice is conservative. While I would like to emphasize that this practice is perfectly legitimate,

it can go wrong when it is coupled with the objection that dynamic semantics fails to be similarly conservative.[25] I take it that dynamic semantics should not be held to this same standard of conservativity, as it directly rejects the idealizing assumptions that give rise to the conservative approach in the first place. While proponents of dynamic semantics do owe explanations as to why the system behaves the way it does, it should not be assumed that non-classical results are unjustified, just because they do not directly explain simple inferences or natural language constructions.

One may worry that this outlook leaves competing semantic theories at loggerheads with no way to resolve disputes. I do not believe this to be the case. Rather, I think that it is overly simplistic to hold radically different semantic theories to precisely the same desiderata.[26] I propose, however, that we may profitable adjudicate disputes between theories by characterizing them in terms of their idealizing assumptions. For example, I have defended dynamic semantics' rejection of Non-Contradiction, but I have in no way espoused any form of dialethism. Rather, I expect Non-Contradiction to hold in any plausible semantics which adopts context fixity. We should only expect it to fail when context fixity is rejected. This is all to say that individuating semantic theories based on their idealizing assumptions should inform, at least in part, what standards they should be held to. This, in turn, opens the door for fresh debate concerning when and why certain idealizing assumptions ought to be made. Such debates commonly occur in science and engineering, and semantics ought to follow suit. Recent debates in semantics place overly narrow focus on cute little puzzles. I believe that explicit discussion of idealizing assumptions may help sort out who is obligated to solve which, and why. I do not take myself to have defended the dynamic semanticist's claim that context fixity should be rejected when theorizing about epistemic modals, but I expect discussion that explicitly recognizes

[25] See Holliday and Mandelkern (2022) for an example of this approach.

[26] I suspect that dynamic semanticists are partly to blame for this, as the extent to which dynamic semantics differs from classical semantics, formally and philosophically, is often underemphasized.

this and other idealizing assumptions will be fruitful. I also suspect that relating theories in terms of their idealizing assumptions might orient them in a way that is less antagonistic than previously thought.

Chapter 4

A Puzzle About Dynamic Belief

Synopsis: I introduce a novel epistemic puzzle that does not occur in static frameworks. The puzzle involves contents which are decisive yet non-committal. Content is decisive with respect to an issue iff an agent who believes that content must be decided on the issue. Content that is non-committal is content that does not commit the believing agent to a particular decision, even if they are decided. Dynamic semantics allows for the unique case where contents can be both decisive yet non-committal, i.e. if an agent believes such content, they must be decided, but the content itself does not commit the believing agent to a particular choice on the issue. This results in an epistemic puzzle whereby undecided agents can update with decisive, non-committal content, and after update, are decided. To solve the puzzle, I appeal to a long forgotten notion of entailment, introduced in Veltman (1996), and largely ignored thereafter, called minimal entailment. After demonstrating how and why minimal entailment allows the dynamic semanticist to solve the puzzle, I recommend that dynamic semanticists more fully embrace the extant practice of utilizing multiple entailment relations within the same semantic framework.

Introduction

Dynamic semantics treats meanings as Context Change Potentials (CCPs) rather than propositions. Content is characterized in terms of update on an information state, and belief is standardly defined as the fixed point of an update. According to this picture, an agent believes some content just in case their information state does not change when updated with that content, i.e. their information state *accepts* that content. Let an information state be the set of possible worlds compatible with the information in the state. When an information state is updated with some atomic formula p, the worlds incompatible with p are removed from the state. Thus, an agent believes p just when p is true at every world in their information state and p is thereby accepted at that state. Importantly, not all updates express propositions, and are thus, not truth-evaluable. Notable among these are "might" updates, which perform a test on an information state. Consider update with 'It might be that p' ($\Diamond p$). According to dynamic semantics, such an update will test the information state for worlds where p is true. If there are any such worlds, the test is passed, and the initial state is returned. Should it fail, all worlds are removed, and the state crashes. Accordingly, any state that contains at least one p world will accept $\Diamond p$, and therefore, any agent who has such a state will believe $\Diamond p$. This is to say that 'might' contents, while neither true nor false, are accepted based on state level properties.

This allows for interesting results that are unavailable in traditional frameworks. For instance, all and only non-absurd states that are undecided about whether p can accept $\Diamond p \wedge \Diamond \neg p$. Once a state decides on p (accepts p or its negation), then the state will no longer support $\Diamond p \wedge \Diamond \neg p$. The content of $\Diamond p \wedge \Diamond \neg p$, however, in no way describes the believing agent as being in an undecided state. With details to follow, let P denote the issue of whether or not p. We can say that $\Diamond p \wedge \Diamond \neg p$ is *indecisive* or *expresses indecision* with respect to P, as all and only states undecided

about P accept this content.[1] This feature distinguishes dynamic semantics from truth-conditional semantics insofar as indecision is characterized by a belief, rather than its absence. Sentences can also be decisive on certain issues. We can say that a formula is decisive with respect to P just in case all states that accept that formula also accept p or its negation. Obvious cases involve updates that express propositions, like p itself. Such contents express decision about the issue of whether or not p by committing to p, as one cannot be committed to p without having made a decision as to whether p. Much more interestingly, dynamic semantics allows for contents that are decisive without committing to a particular decision. A simple example is the negation of the indecisive content above: $\neg(\Diamond p \wedge \Diamond \neg p)$, belief in which ensures that the believing agent either believes p or believes $\neg p$. Such sentences, like the previously considered cases, are decisive, but do not describe the believing agent as being decided. However, unlike straightforwardly decisive contents, they do not express which way the decision went, merely that it has been made. Both agents who believe p and agents who believe $\neg p$, will each believe $\neg(\Diamond p \wedge \Diamond \neg p)$. Agents who are undecided about p will not.

Should an agent who is undecided about P update their information state with $\neg(\Diamond p \wedge \Diamond \neg p)$, it will crash. Once, however, the agent decides whether or not p, the decisive content, $\neg(\Diamond p \wedge \Diamond \neg p)$, is also accepted, and the indecisive content $\Diamond p \wedge \Diamond \neg p$, is no longer acceptable. This acts as an illustration of the failures of persistence that dynamic semantics is known for, while also intuitively capturing the fact that one cannot accept decisive yet non-committal content prior to deciding. Plausibly, one must first decide, and then, in virtue of deciding, one accepts decisive, yet non-committal content. This is what we would like to say, anyhow. The problem is that dynamic semantics recognizes certain non-committal, yet decisive contents that

[1]There are numerous parallels with metaethical expressivism, insofar as certain contents, on some level, *express* certain metal states, like indecision, without describing the agent as being in that state. These parallels, and their limits, are of significant interest, but are not directly relevant to the present argument.

undecided agents can accept. In particular, the content $\neg(\Diamond p \wedge \neg p)$ is decisive in that it is only accepted at states that are committed to p or $\neg p$, but it is also non-committal, as both states that accept p, and states that accept $\neg p$ will accept $\neg(\Diamond p \wedge \neg p)$. However, unlike $\neg(\Diamond p \wedge \Diamond \neg p)$, which will crash any undecided state updated with it, undecided states can be updated with $\neg(\Diamond p \wedge \neg p)$ without crashing. More puzzlingly, despite the fact $\neg(\Diamond p \wedge \neg p)$ is non-committal, any undecided state updated with $\neg(\Diamond p \wedge \neg p)$ will be committed to p; strange behavior for a formula which is non-committal. This result is not only puzzling and unintuitive, it generates scenarios where dynamic semantics makes less than plausible epistemic predictions.

Proponents of dynamic semantics are thus required to explain these results, or rather, how to avoid them. To do so, they must either utilize extant semantic machinery to explain how these implausible results are to be avoided or adopt ancillary epistemic principles. The present chapter pursues the former option and strives to explain how dynamic semanticists can sidestep these puzzling consequences with presently available machinery. The relevant problem cases are difficult to appreciate without an understanding of dynamic semantics, so we proceed as follows. §4.1 rehearses the test semantics and defines affiliated concepts. §4.2 introduces a heretofore, unappreciated puzzle involving commitment and decision in a dynamic environment. §4.3 considers potential strategies for resolving the issue. §4.4 offers a solution which appeals to an underutilized notion of entailment from Veltman (1996). It is then argued that appeal to this notion of entailment is consistent with the common dynamic strategy of utilizing multiple different definitions of entailment and consistency within a single semantics. §4.5 closes by suggesting that this and other notions of entailment should not remain underappreciated.

4.1 Formal Details

4.1.1 Dynamic Semantics

In order to illustrate our puzzle, we consider an off the shelf dynamic semantics, in the form of the propositional fragment of the semantics proposed in Groenendijk and Stokhof (1991b).[2]

Let $\mathcal{L}_1$ be generated by the grammar:

$$\varphi ::= \mathrm{p} \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid \Diamond\varphi$$

Dynamic semantics is intended to capture the manner in which sentences can update a state of information based on their compositional structure. We thus characterize a framework for the way in which sentences can update states of information. This can be captured by an *update system*.

Definition 1: Update System

$\langle \mathcal{L}, C, \cdot[\cdot] \rangle$ is an update system $\Leftrightarrow$ $\mathcal{L}$ is a set of sentences, C is a set of information states, and $\cdot[\cdot]$ is a function which maps sentences of $\mathcal{L}$ to operations on C.

An update system models the manner in which sentences of a given language can change a state of information. Let W be the set of all functions $\mathcal{A} \mapsto \{0, 1\}$. Our update system will be $\langle \mathcal{L}_1, \mathcal{P}(W), \cdot[\cdot] \rangle$. Some information state $s \in \mathcal{P}(W)$ is a set of possible worlds compatible with the information contained within the state. We characterize our semantics for $\mathcal{L}_1$ in terms of operations on $\mathcal{P}(W)$.

Definition 2: Semantic Clauses[3]

[2]See Veltman (1996), Groenendijk and Stokhof (1991a) van der Does et al. (1997), Aloni (2001), Gillies (2004), von Fintel and Gillies (2007), Yalcin (2012b), Yalcin (2015), Willer (2013), and Lennertz (2019) among severaly others, for developments and variations of the test semantics.

[3]The present formulation only considers negation, conjunction, and epistemic possibility operators. Disjunction and quantification are suppressed, as the problem can be formulated in their absence.

1. $s[\mathrm{p}] = \{w \in s | w(\mathrm{p}) = 1\}$

2. $s[\varphi \wedge \psi] = s[\varphi][\psi]$

3. $s[\neg\varphi] = s - s[\varphi]$

4. $s[\Diamond\varphi] = \{w \in s : s[\varphi] \neq \varnothing\}$

Updating a state with an atomic formula takes the intersection of the worlds in the state, and the set of worlds where that atomic is true. Conjunction is consecutive update, and negation takes the difference between the initial state, and what would result by updating the initial state with the non-negated formula. Epistemic possibility (epistemic "might") performs a test, where the initial state is returned if the state is compatible with the prejacent of the modal. Otherwise, the test is "failed" and the absurd state $\varnothing$ is returned.

We can next define what it is for a state to support some information.

Definition 3: Support

$s \vDash \varphi \Leftrightarrow s[\varphi] = s$

A state supports some information if and only if updating the state with that information does not change the state. When a state supports some formula, we can equivalently say that it *accepts* that formula. Famously, three definitions of entailment that are classically equivalent come apart in dynamic semantics. Two receive the lion's share of the attention in the literature concerning epistemic modals.

Definition 4: Update-to-Test Consequence

$\psi_1, ..., \psi_n \Vdash_2 \varphi \Leftrightarrow \forall s, s[\psi_1]...[\psi_n] \vDash_3 \varphi$.

According to update-to-test consequence, a sequence of formula entails another formula φ just in case any state updated with that sequence of formula will yield a state that accepts φ.

Definition 5: Test-to-Test Consequence

$\psi_1, ..., \psi_n \Vdash_3 \varphi \Leftrightarrow \forall s$ s.t. $s \vDash \psi_1, ..., s \vDash \psi_n, s \vDash \varphi$.

According to test-to-test consequence, a set of formula entails some formula φ just in case every state that accepts all of the formula in that set will support φ.

Veltman (1996) introduces the above notions of entailment and labels them with the numerical subscripts above, and each comes with an interdefinable notion of consistency.

Definition 6: Dynamic Consistency

$\psi_1, ..., \psi_n$ is dynamically consistent $\Leftrightarrow \exists s$ s.t. $s[\psi_1]...[\psi_n] \neq \varnothing$

A sequence of formula is dynamically consistent if and only if there exists a state that can sustain update with that sequence without resulting in the absurd state. A sequence is dynamically inconsistent just in case it is not dynamically consistent. Dynamic consistency is interdefinable in with update-to-test consequence and is typically treated as the default notion of entailment in a dynamic setting.[4]

Test-to-test consequence has a similarly interdefinable notion of consistency commonly called *coherence*.

Definition 7: Coherence

$\psi_1, ..., \psi_n$ is coherent $\Leftrightarrow \exists s \neq \varnothing$, s.t. $s \vDash \psi_1, ..., s \vDash \psi_n$

A sequence is coherent just in case there is a state that supports each update individually. It is incoherent otherwise. Like its associated notion of entailment, coherence is not order-sensitive in the way that dynamic consistency is. The upshot is that dynamic semantics allows for several distinct entailment relations corresponding to different "styles" of dynamic inference.[5] Update-to-test consequence and test-to-test consequence, in addition to their associated notions of consistency, are regularly ap-

[4]See Willer (2013). Willer (2015), and Lennertz (2019) for dynamic accounts that appeal to dynamic inconsistency. See Mandelkern (2020) for a critical discussion of dynamic semantics wherein dynamic consistency is taken to be the operative notion of consistency.

[5]Use of the terms"styles" is taken directly from van Benthem (1996).

pealed to in order to explain the infelicity of certain utterances.[6]

A few more definitions will be also useful.

Definition 8: Distributivity

$[\varphi]$ is distributive $\Leftrightarrow$ for all $s, s[\varphi] = \bigcup\{\{w\}[\varphi] | w \in s\}$

An update is distributive just in case updating a state with the formula is equivalent to the union of the individual updates of the singleton set of each world in that state, (van Benthem, 1989, 364). All contents that do not include the $\Diamond$ operator will be distributive. Distributive update can be characterized by the union of an information state and a fixed set of worlds, and thus, updates that are distributive can be said to express propositions. Updates that are not distributive cannot be associated with a particular proposition; updating with such formula will not always intersect with a fixed set of worlds, since the update is sensitive to the state it updates.

While dynamic semantics does not cash out meanings in terms of propositions, propositions can be recovered from certain updates. It will sometimes be useful to refer to propositions directly. Since only formula with distributive updates express propositions, we appeal to the following definition.

Definition 9: Proposition

For any φ that is distributive, $[\![\varphi]\!] =_{\text{def}} \{w \in W | \{w\}[\varphi] \neq \varnothing\}$

Distributive contents can always be associated with a set of possible worlds or *proposition*. $[\![\cdot]\!]$ allows us to refer to this proposition directly. Crucially, $[\![\cdot]\!]$ is only defined for distributive contents.

4.1.2 Doxastic Notions

With our semantics in hand, we now connect it with the appropriate doxastic machinery, beginning with belief. In traditional static frameworks, belief contents will

[6] Willer (2013) for instance appeals not only to dynamic inconsistency, but to incoherence, to explain the infelicity of embedded epistemic contradictions.

be propositions, but this won't be so for dynamic theories. Instead, to have a belief is to have an information state that supports the content of that belief.

Definition 10: Belief

An agent a believes some content $\varphi \Leftrightarrow I_a[\varphi] = I_a$

According to this account of belief, an agent holds a belief just in case update with the contents of that belief will not change the agent's information state. This is to say that the information contained in the update is already captured by the state, and thus, subsequent update has no effect. This can be equivalently defined in terms of support.

Next, we consider the notion of an *issue*. An issue is a subject matter, or what a sentence is about. There are numerous, sophisticated ways in which we might define an issue, but for our purposes, we can follow the characterization provided in Lewis (1988) and treat an issue as a partition on logical space.

Definition 11: Issue

Q is an issue $\Leftrightarrow$ Q is a partition of W, $P \mid R$ s.t. for some proposition $[\![\varphi]\!]$, $P = [\![\varphi]\!]$ and $R = W - [\![\varphi]\!]$.

An issue Q is a partition of logical space that divides all possible worlds into two cells.[7] One cell will include all worlds where that proposition is true, and the other will include all worlds where that proposition is false. For every proposition, there will be an issue associated with that proposition. We can call this the *corresponding issue*. Any given issue will similarly have a pair of *corresponding propositions*. There are more fine-grained characterizations of issues in the literature, but these are not required to generate our puzzle.[8] For our purposes, the crucial feature of issues is that sentences and agents can be decided on certain issues.

[7] We can view an issue as a set of subsets of worlds, with each subset corresponding to one cell of the partition.

[8] Most famously, see **?**.

Definition 12: Decision

φ is decisive with respect to an issue $Q \Leftrightarrow$ for all $s \vDash \varphi$ there exists some cell $P \in Q$ s.t.:

- $s \subseteq P$, and
- For all cells $R \in Q$, if $s \subseteq R$ then $R = P$.

A formula is decisive with respect to an issue just in case every state that supports that formula is in exactly one of the two cells of the partition of that issue.[9] For instance, let A be the issue whose corresponding propositions are a and $\neg a$. Both a and $\neg a$ are decisive with respect to the issue A, because any state that supports either formula will inhabit, at most, one cell of A. More generally, any formula that expresses a proposition will be decisive with respect to that proposition's corresponding issue. Notice that the formula $\neg(\Diamond a \wedge \Diamond \neg a)$ is decisive. However, unlike a where every s s.t. $s \vDash p$ falls within the cell that corresponds with $[\![p]\!]$, the states that support $\neg(\Diamond a \wedge \Diamond \neg a)$ can also fall within the cell that corresponds with $[\![\neg p]\!]$. The formula remains decisive because each state falls under exactly one cell.

In a traditional semantics, the notion of decision would not be interestingly distinct from belief, as every formula expresses a proposition, and is, thereby, decisive with respect to that proposition's corresponding issue. This is not the case in dynamic semantics, where formula with non-distributive updates do not express propositions, and thereby behave uniquely with respect to decision. For instance, $\Diamond a$ is not decisive with respect to A. In fact, $\Diamond a$ is not decisive with respect to any non-trivial issue.[10] Dynamic semantics differs from more classical semantics in that it allows for this to be the case. It also allows for novel behavior involving commitment.

[9] If a state/agent supports/believes the content of some formula that is decided with respect to an issue, we can say, derivatively, that that state/agent is decided on that issue.

[10] Let the trivial issue be the partition of W with one cell equal to W, and the other equal to $\varnothing$. Every sentence is decisive with respect to the trivial issue.

Definition 13: Commitment

φ is committal with respect to an issue $Q \Leftrightarrow$ there is exactly one cell $P \in Q$ s.t. for all $s \vDash \varphi$, $s \subseteq P$.

A formula is committal with respect to an issue, if and only if every state that supports that formula is a subset of one and the same cell of that issue.[11] Once again, a and $\neg a$ are both committal with respect to A. $\Diamond a$ is not committal with respect to A, or any other non-trivial issue. As before, all distributive updates are committal with respect to their corresponding issue, but non-distributive updates behave differently.

If φ is committal, then φ is decisive. However, of particular interest is the fact that certain formula are decisive yet non-committal. For instance, $\neg(\Diamond p \wedge \Diamond \neg p)$, is decisive with respect to P, as only agents who believe p or $\neg p$ will believe $\neg(\Diamond p \wedge \Diamond \neg p)$. Despite being decisive, $\neg(\Diamond p \wedge \Diamond \neg p)$ is non-committal. While any agent who believes $\neg(\Diamond p \wedge \Diamond \neg p)$ will be committed to either p or $\neg p$, $\neg(\Diamond p \wedge \Diamond \neg p)$ does not determine which. Rather, $\neg(\Diamond p \wedge \Diamond \neg p)$ merely expresses that a decision has been made, without expressing which way the decision went. The presence of decisive, yet non-committal content is not present in classical frameworks that deal directly in terms of propositions. Certain decisive, yet non-committal contents yield putatively bizarre results in otherwise benign circumstances.

4.2 A Puzzle About Dynamic Belief

George, Harry, and Isadora have different positions on issue A. Harry believes a, and Isadora believes $\neg a$. George is undecided. George has absolute trust in both Harry and Isadora, and strives to believe everything that Harry and Isadora do. George remains undecided about A precisely because Harry and Isadora disagree. Should one of them be swayed, George would follow suit. In addition, Harry and Isadora

[11] We can say that derivatively, states/agents are committed on an issue if and only if they believe some formula which is committal.

agree on everything else other than A. Since Harry and Isadora are decided, each believes $\neg(\Diamond a \wedge \Diamond\neg a)$. George, on the other hand believes $\Diamond a \wedge \neg\Diamond a$ because he is undecided. Since George would like to believe everything that Harry and Isadora believe, he is frustrated by the fact that they disagree. In addition, since Harry and Isadora agree on every other issue, George has no reason to trust either's opinion over the other. He thus abstains from believing a as well as $\neg a$ because to believe one would be to disbelieve the other. However, both Isadora and Harry believe the decisive, yet non-committal content $\neg(\Diamond a \wedge \Diamond\neg a)$. Can George join the in this belief? He cannot. Despite the fact that both disagreeing agents believe $\neg(\Diamond a \wedge \Diamond\neg a)$, George cannot hold this belief without first believing a or $\neg a$; he can't be decided without making a decision. Were he to update his undecided state with $\neg(\Diamond a \wedge \Diamond\neg a)$, his state would crash.

So far, so good. George is rightly frustrated not only in believing a or $\neg a$ but also in believing $\neg(\Diamond a \wedge \Diamond\neg a)$, and this is exactly what we should expect. The problem, however, is that there are other decisive, yet non-committal contents that both Harry and Isadora believe, and George can also believe. Consider the content $\neg(\Diamond a \wedge \neg a)$.[12] $\neg(\Diamond a \wedge \neg a)$ is decisive because all and only states that support a or $\neg a$ will support his content. In addition, $\neg(\Diamond a \wedge \neg a)$ is non-committal, because this content itself neither entails a nor $\neg a$. We should expect George to be frustrated in coming to accept $\neg(\Diamond a \wedge \neg a)$ for the same reasons that he was frustrated in believing other decisive, yet not committal content. This is not what we observe. If an undecided state is updated with $\neg(\Diamond a \wedge \Diamond\neg a)$ the state is guaranteed to crash, but this is not the case for $\neg(\Diamond a \wedge \neg a)$. George can update with $\neg(\Diamond a \wedge \neg a)$, without crashing, and is not discouraged from doing so because $\neg(\Diamond a \wedge \neg a)$ is believed by both Harry and Isadora. Puzzlingly, despite the fact that $\neg(\Diamond a \wedge \neg a)$ is decisive, and non-committal, should George, or any other undecided agent update with $\neg(\Diamond a \wedge \neg a)$, they will subsequently be committed

[12] This is just the negation of a left modalized epistemic contradiction.

to and believe a! In plain English, this means that in dynamic semantics, there are decisive, yet non-committal contents that are not only acceptable by undecided agents, but updating with such contents forces a particular commitment in undecided agents, yielding them both committed and decided. Consequently, since George wants to believe everything that Harry and Isadora do, and both Harry and Isadora believe $\neg(\Diamond a \wedge \neg a)$ George will too. Accordingly, George updates his information state with $\neg(\Diamond a \wedge \neg a)$ and as a consequence, believes a and thereby, disagrees with Isadora, despite the fact that he updated his information state with something she, herself, believes. This result strikes as fantastically puzzling. George's aim was to believe as much as he could that was shared by both other agents, and so he adopted a belief that they both share. In addition, the belief he adopted is consistent with and supported by both competing positions, and yet, in adopting this belief, George now disagrees with Isadora. In short, George adopted a belief held by Isadora, and in so doing, winds up disagreeing with her.

I take it that this result is undesirable, unintended, and unexpected. In addition, the puzzle can be generated with any propositional formula substituted for a. The semantic machinery in the previous section is perfectly capable of explaining why George should abstain from believing either a or $\neg a$. Each update is inconsistent with the information of one of his interlocutors. Dynamic semantics is similarly well-situated to explain why George cannot join Harry and Isadora in believing $\neg(\Diamond p \wedge \Diamond\neg p)$, since it is inconsistent with George's information. However, no such explanation is available for $\neg(\Diamond p \wedge \neg p)$. Both of his interlocutors believe it, and it is perfectly consistent with his information. What we need, and apparently lack, is an explanation for why George should abstain from believing $\neg(\Diamond p \wedge \neg p)$. The next section considers the options.

4.3 Potential Strategies

Should we wish to defend the dynamic approach, but avoid this odd result, two broad strategies suggest themselves. The first is to stipulate some epistemic principle, over and above the semantics, which explains why George should not adopt $\neg(\Diamond a \wedge \neg a)$. This option strikes as profoundly concessionary. Dynamic semantics is designed to model thought and talk which deals directly with uncertainty and is often utilized to model things like disagreements. More generally, a semantics is obligated to precisely capture the kinds of relationships between pieces of information that our puzzle is concerned with. A more promising strategy is to search for a solution based on the extant semantic and epistemic machinery or change that machinery in order to avoid the puzzle.

4.3.1 Adjusting Indecision

The simplest characterization of belief in dynamic frameworks is to take belief as the fixed point of update. There are proposals in the literature that add layers to this account may be able to diffuse the puzzle. For example, the version of dynamic semantics proposed in Willer (2013) has added structure that alters the update effects of epistemic modals. For Willer, information states are not sets of worlds, but sets of sets of worlds. Call each set of worlds within a state a *substate*. When an update is performed on a state, updates are applied to each substate, and then substates are aggregated. The result is that epistemic modals do not simply test for compatibility with the prejacent. Rather, a state will only support Might p on condition that every substate contains at least one p world. This is intended to capture the idea that sentences like Might p express more than mere compatibility with p, but that p is taken seriously, or matters, in inquiry, (Willer, 2013, 56). Accordingly, an agent's information state will only support Might$p$ if that agent is disposed to take p seriously

as a possibility.

Under Willer's proposal, $\Diamond a \wedge \Diamond \neg a$ is not supported by every undecided state. Rather, it is only supported by states that take both a and $\neg a$ seriously. While such an approach would substantially change what it is for an agent to be undecided, it does nothing to help with our puzzle involving George, Harry and Isadora. George takes both a and $\neg a$ seriously and believes $\Diamond a \wedge \Diamond \neg a$. Similarly, under Willer's account, both George and Isadora believe $\neg(\Diamond a \wedge \neg a)$. In addition, $\Diamond a \wedge \Diamond \neg a$ is still undecisive, and $\neg(\Diamond a \wedge \neg a)$ is still decisive, but non-committal. Accordingly, Willer's account may fair better than the simple account in situations where George takes one side more seriously than the other, but this condition does not hold in our puzzle. Under these circumstances, Willer's account makes predictions that are no different from the simple account.

Yalcin (2018) offers an account of belief where belief is *question sensitive*. Belief is relativized to questions or *resolutions* of logical space, (Yalcin, 2018, 30).[13] The intuitive idea is that agents can sometimes fail to recognize certain distinctions in the logical space, i.e. the agent can fail to ask certain questions. When an agent does recognize a distinction, her logical space will represent at a higher resolution, individuating possibilities that it previously did not. The account is intended to solve various puzzles concerning logical omniscience, but, for our purposes, the view allows for a distinction between indecision about a, and believing $\Diamond a \wedge \Diamond \neg a$. For instance, if an agent has never considered the issue of whether or not a, their logical space will not individuate a worlds from $\neg a$ worlds. The same goes for agents that lack some relevant concept crucial to understanding a, (Yalcin, 2018, 34). The immediate consequence, once again, is that all agents who are undecided about a are not guaranteed to believe $\Diamond a \wedge \Diamond \neg a$, and our account of indecision must be altered accordingly. As with Willer's account, however, this does nothing to solve the puzzle involving George, Harry, and

[13] A similar story is provided in Yalcin (2011).

Isadora. Every agent's logical space is sensitive to A and so the puzzle remains.

Thus, despite the many virtues of Willer and Yalcin's proposals, they offer no assistance with respect to our particular puzzle.

4.3.2 Inversion Cases

Careful readers may have noticed that our puzzle only emphasizes one belief that is decisive yet non-committal, namely, $\neg(\Diamond a \wedge \neg a)$. There are however, several at play, and they are not all equivalent. $\neg(\Diamond a \wedge \neg a)$ has a sinister twin, in $\neg(\Diamond \neg a \wedge a)$ where the negations are swapped. Like its brother, $\neg(\Diamond \neg a \wedge a)$ is decisive, but non-committal, and is believed by both Harry and Isadora. However, should an undecided agent update with $\neg(\Diamond \neg a \wedge a)$, they will subsequently believe $\neg a$. Call this an *inversion case*.

Inversion cases complicate the puzzle. Given that George wants to believe everything that Harry and Isadora believe, and since Harry and Isadora both believe $\neg(\Diamond a \wedge \neg a)$ and its inversion, $\neg(\Diamond \neg a \wedge a)$, George is inclined to follow suit. However, if George updates with $\neg(\Diamond a \wedge \neg a)$ he will subsequently believe a, but if George instead updates with $\neg(\Diamond \neg a \wedge a)$ he will subsequently believe $\neg a$. We may optimistically expect this tension to resolve the puzzle. After all, George abstains from believing a precisely because it is inconsistent with $\neg a$. Unfortunately, this sort of explanation is unavailable for inversion cases. George is unwilling to update with a because a is inconsistent with $\neg a$. Similarly, $\neg a$ is off limits, because it is inconsistent with a. This is not the case with $\neg(\Diamond a \wedge \neg a)$ and $\neg(\Diamond \neg a \wedge a)$, since any state updated with one will immediately support the other; any decided state supports both. In addition, any sequence, regardless of order, of $\neg(\Diamond a \wedge \neg a)$ and its inversion will be both dynamically consistent as well as coherent. Thus, there is no obvious semantic explanation for why George shouldn't believe both.

Perhaps more puzzlingly, should George decide to update with $\neg(\Diamond a \wedge \neg a)$ his state

will subsequently support a, as well as the inversion, $\neg(\Diamond\neg a \wedge a)$. However, should George instead update with $\neg(\Diamond\neg a \wedge a)$ his subsequent state will support $\neg a$ as well as $\neg(\Diamond a \wedge \neg a)$. The result is that George's choice with respect to which update he chooses first, will determine whether he believes a or $\neg a$. If he updates with either, he'll believe the other, but the chosen order of updates fundamentally changes the nature of which way he will decide. This, again, is highly bizarre. Epistemically speaking, the choice seems ill-posed, and one may suspect that the choice itself is an artifact of dynamic meaning that is in tension with any traditional approach to epistemology. Any answer which explains or embraces either choice strikes, at least intuitively, as epistemically misguided. In any event, the presence of inversion cases certainly doesn't solve the puzzle and seems to underscore the choice point laid out at the beginning of this section. Indeed, it seems that the semantics itself (in its extant form or modified) will need to provide an explanation for how this choice is to be avoided, or else the dynamic semanticist must concede that substantial epistemological revision is required if we are to appeal to dynamic content in our epistemic theorizing. This would be an especially bitter pill to swallow given how much epistemic machinery is already baked in the semantics.

4.3.3 Coherence

As was mentioned in §4.2, dynamic semantics features numerous distinct notions of entailment and consistency. Of the options, dynamic consequence and consistency ($\Vdash_2$ above) are often taken as the default. Dynamic semanticists typically associated dynamic inconsistency with disagreement and contradiction, and alternative notions of entailment often go unmentioned.

Certain crucial behaviors involving epistemic modals are not captured by dynamic inconsistency.[14] The standard play, for dynamic semanticists, is that when update-

[14]For instance, epistemic contradictions like $\Diamond p \wedge \neg p$ where the left conjunct is modalized, are

to-test consequence and its affiliated notions fail, appeal to test-to-test consequence ($\Vdash_3$ above). Test-to-test consequence's indefinable notion of consistency is commonly referred to as coherence. Incoherence, typically plays second fiddle to dynamic inconsistency, and picks up the scraps that it leaves behind. It might be hoped that this strategy can sort out our puzzle about belief, but unfortunately, this is not the case. Singleton sequences of formula that are decisive but non-committal are perfectly coherent. In addition, sequences of such formula and their inversion cases are perfectly coherent as well. More generally, any state that is undecided with respect to P will believe $\Diamond p \wedge \Diamond \neg p$. Any sequence that contains this formula as well as some formula that is decisive with respect to the same issue will be incoherent. This is a poor predictor of acceptance and rejection, however, as any new propositional information will always be incoherent with some belief in a state that is undecided with respect to that proposition's corresponding issue. This does not prevent agents from uptaking new beliefs, nor should it, as coherence is not intended to play this role. Thus, the standard dynamic strategy of appealing to coherence is of no help, and we'll help to look elsewhere for a solution to our puzzle.

4.4 Minimal Consequence

It was mentioned that there are three distinct notions of entailment, despite the fact that only two receive all of the attention. The third, so far as I am aware, is not mentioned in the literature on epistemic modals after its introduction in Veltman (1996). Accordingly, it has no catchy name, and its interdefinable notion of consistency has never been formally defined.[15] In the absence of a better alternative, call this notion of consequence *minimal update-to-test consequence*, or *minimal consequence* for short.

dynamically consistent, despite their infelicity.

[15]It also remains unmentioned in the chapter of van Benthem (1996) that discusses the numerous entailment relations available to dynamic systems.

Definition 14: Minimal Update-to-Test Consequence

$\psi_1, ..., \psi_n \Vdash_1 \varphi \Leftrightarrow W[\psi_1]...[\psi_n] \vDash \varphi$.

According to minimal consequence ($\Vdash_1$), a sequence of formula minimally entails a formula φ just in case updating the minimal state W with that sequence will yield a state that accepts φ. Rather than defining consequence in terms of a universally quantified statement, it only considers the effect of a sequence of updates on the state with minimal information, W. The related notion of consistency can be defined as follows.

Definition 15: Minimal Consistency

$\psi_1, ..., \psi_n$ is minimally consistent $\Leftrightarrow W[\psi_1], ..., [\psi_n] \neq \varnothing$

A sequence of updates is minimally consistent if and only if that string of updates does not crash the minimal state, W. Heretofore, theorists exploring dynamic behavior involving epistemic modals have seemingly failed to discover any practical application of minimal consequence. This is, perhaps, unsurprising, given its somewhat odd characterization. Minimal consequence only concerns the behavior of updates on the maximally undecided state and lacks the familiar appeal to universal quantification characteristic of most entailment relations. Dynamic semantics is often marshaled to solve puzzles involving thought and talk that concerns various information states, and we are rarely interested in particular focus on the highly idealized minimal state. Notice, however, that our puzzle is different. The puzzle arises from the discrepancy in behavior of decisive yet non-committal updates on undecided versus decided states. When an undecided agent updates with $\neg(\Diamond p \wedge \neg p)$ their updated state will support p. Decided agents, even those who believe $\neg p$ will still support $\neg(\Diamond p \wedge \neg p)$ and so update has no effect. Accordingly, despite the fact that $\neg(\Diamond p \wedge \neg p)$ does not dynamically entail p, any undecided state that sustains this update will support p. In order to solve our puzzle, we need a notion of consequence that solely focuses on the behavior

of updates on undecided states. It's time to get minimal consequence off the bench and into the game.

Our puzzle requires an explanation as to why George should not update with $\neg(\Diamond a \wedge \neg a)$. With minimal consequence in hand, I can now deliver. While $\neg(\Diamond a \wedge \neg a)$ does not dynamically entail a, it does minimally entail a. Similarly, its inversion, $\neg(\Diamond \neg a \wedge a)$ minimally entails $\neg a$. Of course, a is inconsistent with Isadora's information, and $\neg a$ is inconsistent with Harry's. Since George does not want to disagree with either, he abstains from believing $\neg(\Diamond a \wedge \neg a)$ and its inversion, since the minimal entailments of these formula are inconsistent with either Isadora's beliefs, or Harry's beliefs. More generally, van Benthem associates different notions of entailment available in dynamic frameworks with different styles of inference. Minimal entailment captures a style of inference sensitive to transitions from undecided to decided states. Crucially, the styles of inference captured by update-to-test and test-to-test consequence are not sensitive to these transitions. Minimal consequence, however, is, and is able capture the style of reasoning available to George, that would cause him to abstain from update with decisive, yet non-committal contents.

While minimal entailment can provide a solution to our puzzle, it may be less clear why minimal entailment is the right notion for the job. The idea is something like this. Our trouble arose because, despite the fact that $\neg(\Diamond a \wedge \neg a)$ does not entail a, it has the odd consequence of ensuring that any state undecided about A will support a if updated with $\neg(\Diamond a \wedge \neg a)$. Update-to-test consequence fails to capture this result, but minimal consequence does. It is able to do this in virtue of the fact that the minimal state, W is undecided with respect to every non-trivial issue. Accordingly, minimal consequence is able to capture the results that certain updates uniquely produce on undecided states. This suggest that minimal entailment may be of use in other puzzles that specifically concern indecision in dynamic semantics or require this style of inference.

Despite its brevity, this explanation is noteworthy for two reasons. The first is that it fundamentally appeals to minimal consequence, which has, heretofore, remained underutilized. The second feature is that it does not merely appeal to minimal consequence, but also appeals to dynamic inconsistency. Minimal consequence allows George to determine what he would believe, should he update with $\neg(\Diamond a \wedge \neg a)$ while update-to-test consequence, by way of dynamic inconsistency, explains the disagreement that George is striving to avoid. Thus, the dynamic explanation of why George should abstain from performing this update appeals to, not one, but two distinct entailment relations. I take the second point to be more crucial than the first. In having multiple distinct consequence relations, dynamic semantics recognizes several semantic/inferential relations that can hold between pieces of information. The key, however, is that these inferential relations do not exist in a vacuum, and explanations of more complex actions and decisions may be the result of their interactions.

Critics and proponents of dynamic semantics often behave as if update-to-test consequence enjoys the position of the notion of entailment *de jure*, we should neither assume, nor passively behave as if this is the case. Indeed, it was mentioned previously that dynamic semantics often appeals to both dynamic inconsistency and incoherence to predict the infelicity of different utterances. This is to say that it is already in line with dynamic practice to appeal to these distinct notions when operating in a single dynamic framework. These distinct notions of entailment model different styles of inference, and by appealing to each, we expand our capacity to model the various kinds of inferences that agents can make concerning what they should believe, say, and update with. While this strategy has been embraced by earlier work on dynamic semantics, it does not seem to manifest itself in the same full-throated manner when dynamic semantics concerns epistemic modals. I do not think that this should be the case, and I think semanticists in the dynamic tradition should more fully embrace and explore the myriad notions of entailment that dynamic semantics makes available.

4.5 Concluding Remarks

In this chapter, I have argued that dynamic semantics faces certain puzzles involving sentences which were decisive yet non-committal. After considering potential solutions, I argue that the underappreciated notion of minimal entailment can succinctly solve the puzzle. I suggest that minimal entailment should not remain underappreciated, and that more generally, dynamic semanticists should make open appeal to numerous consequence relations when theorizing in a dynamic framework.

Chapter 5

Conclusion

In the preceding chapters I've considered puzzles, challenges, and objections to dynamic semantics and have done my best to address them. Each challenge is unique to dynamic semantics and is not faced by competing truth-conditional theories. I take each of these to underscore the theme of the dissertation, which is that dynamic semantics is radically different from truth-conditional semantics, both formally and philosophically. Genuine adoption and defense of the program cannot amount to observations of the program's success with respect to a handful of narrow puzzles. Rather, the differences above must be embraced, and in some cases explained. Cases where either are difficult will serve as a guide to the contours of a more successful and complete manifestation of the program, if such there be.

Appendix i

Proofs for Chapter 3

Theorem 1: $(\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p)$ is dynamically consistent.

Let w be world s.t. $w(p) = 0$ and let w' be a world s.t. $w'(p) = 1$. Let s be a state s.t. $w \in s$ and $w' \in s$.

$$\begin{aligned}
s[(\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p)] &= s[(\Diamond p \wedge \neg p)][\neg(\Diamond p \wedge \neg p)] \\
&= s[\Diamond p][\neg p][\neg(\Diamond p \wedge \neg p)] \\
&= \{w \in s : s[p] \neq \varnothing\}[\neg p][\neg(\Diamond p \wedge \neg p)] \\
&= s[\neg p][\neg(\Diamond p \wedge \neg p)] \\
&= s - s[p][\neg(\Diamond p \wedge \neg p)] \\
&= \{w'\}[\neg(\Diamond p \wedge \neg p)] \\
&= \{w'\} - \{w'\}[(\Diamond p \wedge \neg p)] \\
&= \{w'\} - \{w'\}[(\Diamond p][\neg p] \\
&= \{w'\} - \{w \in \{w'\} : \{w'\}[p] \neq \varnothing\}[\neg p] \\
&= \{w'\} - \varnothing[\neg p] \\
&= \{w'\} - \varnothing - \varnothing[p] \\
&= \{w'\} - \varnothing - \varnothing \\
&= \{w'\} - \varnothing \\
&= \{w'\}
\end{aligned}$$

$\{w'\} \neq \varnothing$ so $(\Diamond p \wedge \neg p) \wedge \neg(\Diamond p \wedge \neg p)$ is dynamically consistent.

■

Theorem 2: $[\varphi]$ is idempotent $\Leftrightarrow \varphi \Vdash \varphi$.

Left to right: Suppose, for conditional proof, that $\varphi \nVdash \varphi$. It follows from the definition of dynamic consequence that it is not the case that for all s, $s[\varphi] \vDash \varphi$. Suppose, for reductio, that $[\varphi]$ is idempotent. It follows from the definition of idempotence that for all s, $s[\varphi] = s[\varphi][\varphi]$. It follows from the definition of support, that for all $s, s[\varphi] \vDash \varphi$. By reductio, $[\varphi]$ is not idempotent. By conditional proof, we have that $\varphi \nVdash \varphi$ only if $[\varphi]$ is not idempotent. By contraposition, we have $[\varphi]$ is idempotent only if $\varphi \Vdash \varphi$.

Right to left: Suppose, for conditional proof, that $[\varphi]$ is not idempotent. It follows from the definition of idempotence that there exists an s s.t. $s[\varphi] \neq s[\varphi][\varphi]$. Thus, there exists a state, s s.t. $s[\varphi] \neq s$. It follows from the definition of entailment that $\varphi \nVdash \varphi$. Therefore, by conditional proof $[\varphi]$ is not idempotent only if $\varphi \nVdash \varphi$. By contraposition, $\varphi \Vdash \varphi$ only if $[\varphi]$ is idempotent.

■

Theorem 3: $s \vDash \varphi$ only if $s[\neg\varphi] = \varnothing$.

Suppose, for conditional proof, that $s \vDash \varphi$. It follows from the definition of support that $s[\varphi] = s$. Now consider $s[\neg\varphi]$.

$$\begin{aligned} s[\neg\varphi] &= s - s[\varphi] \\ &= s - s \\ &= \varnothing \end{aligned}$$

■

Theorem 4: If $\varphi \wedge \neg\varphi$ is consistent, then $[\varphi]$ is not idempotent.

Suppose, for conditional proof, that $\varphi \wedge \neg\varphi$ is consistent. It follows from the definition of consistency that there exists a state s' s.t. $s'[\varphi \wedge \neg\varphi] \neq \varnothing$. It follows from the

clause for conjunction that $s'[\varphi][\neg\varphi] \neq \varnothing$. Now suppose, for reductio, that $[\varphi]$ is idempotent. It follows from Theorem 2 that $\varphi \Vdash \varphi$. By the definition of entailment, it follows that for all $s, s[\varphi] \vDash \varphi$. Therefore, $s'[\varphi] \vDash \varphi$. From Theorem 3, it follows that $s'[\varphi][\neg\varphi] = \varnothing$, but this contradicts $s'[\varphi][\neg\varphi] \neq \varnothing$ established above. Therefore, $[\varphi]$ is not idempotent.

■

Theorem 5: If φ is of the form $\Diamond\psi$ then $[\varphi]$ is idempotent.

Suppose φ is of the form $\Diamond\psi$. Suppose, for reductio, that there is some state s s.t. $s[\varphi] \neq s[\varphi][\varphi]$. Because $[\cdot]$ is eliminative, if $s[\varphi] = \varnothing$ then $s[\varphi] = s[\varphi][\varphi]$. By modus tollens $s[\varphi] \neq \varnothing$. However, since φ is of the form $\Diamond\psi$, if $s[\varphi] \neq \varnothing$ then $s[\varphi] = s$, and if $s[\varphi] = s$, then $s[\varphi] = s[\varphi][\varphi]$. Therefore, $s[\varphi] = s[\varphi][\varphi]$, but this contradicts the assumption for reductio. Therefore, for all s, $s[\varphi] = s[\varphi][\varphi]$ and therefore, $[\varphi]$ is idempotent.

■

Theorem 6: If $s \vDash \neg p$ then $s \vDash \neg(\Diamond p \wedge \neg p)$

Proof: Suppose $s \vDash \neg p$

$$\begin{aligned} s[\neg(\Diamond p \wedge \neg p)] &= s - s[\Diamond p \wedge \neg p] \\ &= s - s[\Diamond p][\neg p] \\ &= s - \varnothing[\neg p] \\ &= s - \varnothing \\ &= s \end{aligned}$$

■

Theorem 7: If φ is of the form $\neg\Diamond\psi$ then $[\varphi]$ is idempotent.

Proof: Suppose, for conditional proof, that φ is of the form $\neg\Diamond\psi$.

Suppose, for reductio, that $[\varphi]$ is not idempotent. It follows that there exists a state s s.t.

$$\begin{aligned} s[\neg\Diamond\psi] &\neq s[\neg\Diamond\psi]s[\neg\Diamond\psi] \\ s - s[\Diamond\psi] &\neq s - s[\Diamond\psi][\neg\Diamond\psi] \\ s - \{w \in s : s[\psi] \neq \varnothing\} &\neq s - \{w \in s : s[\psi] \neq \varnothing\}[\neg\Diamond\psi] \end{aligned}$$

If $\{w \in s : s[\psi] \neq \varnothing\} = \varnothing$, then

$$\begin{aligned} s - \varnothing &\neq s - \varnothing[\neg\Diamond\psi] \\ s - \varnothing &\neq s - \varnothing \end{aligned}$$

Therefore, $\{w \in s : s[\psi] \neq \varnothing\} \neq \varnothing$. If $s = \varnothing$ then

$$\begin{aligned} \varnothing[\neg\Diamond\psi] &\neq \varnothing[\neg\Diamond\psi][\neg\Diamond\psi] \\ \varnothing &\neq \varnothing[\neg\Diamond\psi] \\ \varnothing &\neq \varnothing \end{aligned}$$

Therefore, $s \neq \varnothing$. Since $s[\Diamond\psi] = s$ or $s[\Diamond\psi] = \varnothing$ and $\{w \in s : s[\psi] \neq \varnothing\} \neq \varnothing$, it

follows that $s[\Diamond\psi] = s$. Therefore,

$$\begin{aligned} s[\Diamond\psi][\neg\Diamond\psi] &\neq s[\Diamond\psi][\neg\Diamond\psi]s[\neg\Diamond\psi] \\ s[\Diamond\psi] - s[\Diamond\psi] &\neq s[\Diamond\psi] - s[\Diamond\psi][\neg\Diamond\psi] \\ \varnothing &\neq \varnothing[\neg\Diamond\psi] \\ \varnothing &\neq \varnothing \end{aligned}$$

Therefore, by reductio, $[\varphi]$ is idempotent.

■

Bibliography

Aloni, M. (2001) "Quantification Under Conceptual Covers". (Unpublished doctoral dissertation). University of Amsterdam.

Beaver, D. (2001) "Presupposition and Assertion in Dynamic Semantics". CSLI Publications.

Dever, J. (2013). "The Revenge of The Semantics-Pragmatics Distinction". In *Philosophical Perspectives* vol.27 104-144.

Egan, A. (2007). "Epistemic Modals, Relativism, and Assertion". In *Philosophical Studies* vol.133, 1-22.

Egan, A. Weatherson, B. Hawthorne, J. (2005). "Epistemic Modals in Context". In *Contextualism in Philosophy* Eds. Preyer and Peter. Oxford Univerity Press, 131-169.

von Fintel , K. Gillies, A. 2008. "CIA Leaks". In *The Philosophical Review* Vol. 117, (1) 77-98.

von Fintel , K. Gillies, A. 2007. "An Opinionated Guide to Epistemic Modals". In T. Szabó Gendler and J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 2). Oxford: Oxford University Press.

von Fintel , K. Gillies, A. 2010. "Must... Stay... Strong!". In *Natural Language Semantics* vol.18 (4) 351-383.

von Fintel , K. Gillies, A. 2011. "Might made Right". In *Epistemic Modality* Eds. Egan and Weatherson. Oxford university Press, 108-130.

Frege, G. (1918). "Der Gedanke". Beiträge zur Philosophie des deutschen Idealismus, vol.1, 58–77

Gillies, A. (2004). "Epistemic Conditionals and Conditional Epistemics". In *Nous* vol.38(4): 585-616.

Goldstein, S. (2019). "Generalized Update Semantics". In *Mind* vol.128(515): 795-835.

Groenendijk, J. and M. Stokhof (1991a). "Two Theories of Dynamic Semantics". In *In Logics in AI, Volume 478 of Lectures Notes in Computer Science*. Springer, 55-64.

Groenendijk, J. and M. Stokhof (1991). "Dynamic Predicate Logic". In *Linguistics and Philosophy*. vol.14 (1), 39-100.

Hawthorne, J. (2007). "Eavesdroppers and Epistemic Modals". In *Philosophical Issues* vol.17, 92-101.

Heim, I. (1983). "File Change Semantics and the Familiarity Theory of Definiteness". In *Meaning, Use, and Interpretation of Language* Eds. R. Bauerle, C. Schwarze, and A. von Stechow. Walter de Gruyter, 164-190.

Heim, I. (1992). "Presupposition Projection and the Semantics of Attitude verbs". *Journal of Semantics*, vol.9: 183-221.

Holliday, W., Mandelkern, M. (2022). "The orthologic of epistemic modals.". Manuscript, https://arxiv.org/abs/2203.02872

Huvenes, T. (2017). "On Disagreement". In *Routledge Handbook of Epistemic Contextualism* Ed. Ichikawa, J. Routledge. 272-280.

Kamp, H. (1981). "A Theory of Truth and Semantic Representation". In *Formal Methods in the Study of Language* Eds. J. A. Groenendijk, T. Janssen, and M. Stokhof. Mathematish Centrum, 277-322.

Kaplan, D. (1989). "Demonstratives". In *Themes from Kaplan*. Eds. Almog, J. Perry, J. Wettstein, H. Oxford University Press, 481-563.

Karttunen, L. (1972). "*Possible* and *Must*". In *Syntax and Semantics vol.1* Academic Press. 1-20.

Khoo, J. (2015). "Modal Disagreements". *Inquiry* vol.58: 511-534.

Khoo, J. (2017). "The Disagreement Challenge to Contextualism". In *Routledge Handbook of Epistemic Contextualism* Ed. Ichikawa, J. Routledge. 257-271.

Knobe, J. Yalcin, S. (2014). "Epistemic Modals in Context: Experitmental Data". In *Semantics and Pragmatics* vol.7, Article 10 1-21.

Kratzer, A. (1977). "What 'Must' and 'Can' Must and Can Mean". In *Words, Worlds, and Contexts: New Approaches to Word Semantics*, ed. Eikmeyer, and Rieser, Berlin: de Gruyter. 38–74.

Kratzer, A. (1981). "The Notional Category of Modality". In *Linguistics and Philosophy*, vol. 1, 337-355.

Lennertz, B. (2019). "Might-Beliefs and Asymmetric Disagreement". *Synthese* vol.196: 4775–4805.

Lennertz, B. (2020). "Evaluating the Multiple Proposition Strategy". *Ratio* 00: 1–10.

Lewis, D. (1980). "Index, Context, and Content". In *Philosophy and Grammar*. Eds. anger, S., Ōhman, S. Synthese Library, vol 143. Springer.

Lewis, D. (1988). "Statements Partly About Observation". In *Philosophical papers*, vol.17 1-30.

MacFarlane, J. (2003). "Future COntingents and Relative Truth". In *The Philosophical Quarterly* vol.53, 321-336.

MacFarlane, J. (2007). "Relativism and Disagreement". In *Philosophical Studies* vol.132, 17-31.

MacFarlane, J. (2011). "Epistemic Modals are Assessment Sensitive". In *Epistemic Modality* Eds. Egan and Weatherson. Oxford University Press, 144-178.

MacFarlane, J. (2014). *Assessment Sensitivity*. Oxford University Press.

Mandelkern, M. (2019). "Bounded Modality". In *The Philosophical Review* vol.128 (1) 1-61.

Mandelkern, M. (2020). "Dynamic Non-Classicality". In *Australian Journal of Philosophy* vol.98 (2) 382-392.

Moss, S. (2015). "On the Semantics and Pragmatics of Epistemic Vocabulary". In *Semantics and Pragmatics* vol.8, Article 5, 1-81.

Moss, S. (2018). "Probabilistic Knowledge". Oxford University Press.

Nouwen, R. Brasoveanu, A. van Eijck, A. Visser, A. (2016). "Dynamic Semantics". In *The Stanford Encyclopedia of Philosophy* Ed. Zalta. `<https://plato.stanford.edu/entries/dynamic-semantics/>`. Accessed 1/2/2020.

Plunkett, D., Sundell, T. (2013). "Disagreement and the Semantics of Normative and Evaluative Terms". In *Philosopher's Imprint* vol.13(23): 1-37.

Potochnik, A. (2017). "Idealization and The Aims of Science". The University of Chicago press.

Predelli, S. (2005). *Contexts*. Oxford University Press.

Rothschild, J. Yalcin, S. (2016). "Three Notions of Dynamicness in Language". In *Linguistics and Philosophy* vol.39, 333-355.

Rothschild, J. Yalcin, S. (2017). "On the Dynamics of Conversation". In *Nous* vol.52 (2), 24-48.

Schaffer, J. (2011). "Perspective in Taste Predicates and Epistemic Modals". In *Epistemic Modality* Eds. Egan and Weatherson. Oxford university Press, 179-226.

Schlenker, P. (2009). "Local Contexts". In *Syntax and Semantics* vol.3, 1-78.

Stalnaker, R. (1978). "Assertion". In *Syntax and Semantics 9: Pragmatics*, Ed. Peter Cole. 315–332. Academic Press.

Stalnaker, R. (1998). "On the Representation of Context". In *Journal of Logic, Language, and Information* vol.7, 3-19.

van Benthem, J. (1989). "Semantic Parallels in Natural Language and Computation". In Ebbinghaus, H. D., Fernandez-Prida, J., Garrido, M., Lascar, D. and Artalejo, M. R. (eds.) *Logic Colloquium '87: Proceedings of the Colloquium held in Granada, Spain*. Volume 129 of Studies in Logic and the Foundations of Mathematics. 337–375. Amsterdam: North-Holland.

van Benthem, J. (1996). "Exploring Logical Dynamcis". CSLI Publications.

van der Does, J., Groeneveld, W.Veltman, F. (1997). "An Update on "Might". In *Journal of Logic, Language, and Information* 6(4): 361–380.

Veltman, F. (1996). "Defaults in Update Semantics". In *Journal of Philosophical Logic* vol.25, 221-261.

Wedgewood, R. (2002). "The Aim of Belief". In *Philosophical Perspectives* vol.16 267-297.

Willer, M. (2013). "New Dynamics for Epistemic Modality". In *Philosophical Review* vol.122 (1) 45-92.

Willer, M. (2015). "An Update on Epistemic Modals". In *Journal of Philosophical Logic* vol.44 (6) 835-849.

Williams, B. (1973). "Deciding to Believe". In *Problems of the Self*. Cambridge University Press. 136-151.

Yablo, S. (2014). *Aboutness*. Princeton University Press.

Yalcin, S. (2007). "Epistemic Modals". In *Mind* vol.116 (464), 983-1026.

Yalcin, S. (2011). "Nonfactualism About Epistemic Modality". In *Epistemic Modality* Eds. Egan and Weatherson. Oxford university Press, 295-332.

Yalcin, S. (2012a). "Introductory Notes on Dynamic Semantics". In *Routledge Companion to the Philosophy of Language* Eds. Fara and Russell. Routledge, 253-279.

Yalcin, S. (2012b) "Bayesian Expressivism". In *Proceedings of the Aristotlean Society* vol.112 (2), 123–160.

Yalcin, S. (2015) "Epistemic Modality *de re*". In *Ergo* vol.2 (19), 475-527.

Yalcin, S. (2018) "Belief as Question Sensitive". In *Philosophy and Phenomenological Research* XCVII (1), 23-47.

Zardini, E. (2014). "Context and consequence. An intercontextual substructural logic". *Synthese* vol.191: 3473–3500.