

Paradox and Belief

by

Michael Caie

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Philosophy

in the

Graduate Division
of the
University of California, Berkeley

Committee in charge:
Professor John MacFarlane, Co-chair
Professor Michael Martin, Co-chair
Professor Tania Lombrozo

Fall 2011

Paradox and Belief

Copyright 2011
by
Michael Caie

Abstract

Paradox and Belief

by

Michael Caie

Doctor of Philosophy in Philosophy

University of California, Berkeley

Professor John MacFarlane, Co-chair

Professor Michael Martin, Co-chair

At a fairly high level of abstraction, this work is about some ways in which questions about the correct treatment of the semantic paradoxes and questions about the principles of rationality governing doxastic states can be mutually illuminating.

In the first part of the dissertation, I argue that certain treatments of the semantic paradoxes lead to surprising conclusions about the nature of the doxastic states of rational agents. The semantic paradoxes, such as the liar paradox, provide us with good reason to take seriously various non-classical logics. In addition to the semantic paradoxes, there are also paradoxes that show that some extremely plausible principles of rationality governing doxastic states are inconsistent given classical logic. I show how various non-classical responses to the semantic paradoxes provide us with resources sufficient to resolve these paradoxes. In particular, if we allow that certain statements about an agent's doxastic state, e.g., statements about whether an agent believes a proposition ϕ , may give rise to certain failures of classical logic, then we can hold on to all of our plausible principles of doxastic rationality. I use this fact to argue for the conditional claim that *if* one is inclined to reject classical logic in response to the liar paradox, then one should allow that statements about an agent's doxastic state may also give rise to failures of classical logic. Since the antecedent of the conditional is reasonable, and the consequent surprising, the conditional is of interest.

In the second part of the dissertation, I argue that attention to questions about the nature of doxastic rationality can provide us with important insights into the correct treatment of the semantic paradoxes. For any non-classical response to the semantic paradoxes, an important question that arises is: what exactly is the cognitive significance of the non-classical semantic statuses employed by the theory? I argue that our earlier reflections on the normative paradoxes show that the standard ways of answering this question are wrong. Given standard accounts of the cognitive significance of non-classical semantic statuses, we can resurrect our normative paradoxes. What this means is that the standard accounts of non-classical cognitive significance are in conflict with certain fundamental principles of doxastic rationality. I argue that in order to reconcile the account of non-classical cognitive significance with these principles we need to say that the correct rational response to paradoxical propositions, such as that expressed by the liar sentence, is for there to

be a mirroring non-classicality in one's doxastic state. A rational agent, then, will be such that the claim that it believes the proposition expressed by the liar sentence will have the same non-classical status as the proposition expressed by the liar sentence. What emerges is a new picture of the significance of non-classical treatments of the semantic paradoxes.

Contents

Acknowledgments	iii
1 Introductory Remarks	1
2 A Brief Introduction to Theories of Semantic Paradox	5
2.1 The Liar Paradox	5
2.2 Treatments of the Liar Paradox	8
2.2.1 Models	8
2.2.2 Validity	16
2.2.3 Extensions	21
2.3 Going Ahead	26
3 Attitudinal Paradoxes	28
3.1 Belief	28
3.2 Credence	32
3.3 Knowledge	34
3.4 Some Responses	35
3.4.1 Responses to the Paradox of Belief	35
3.4.2 Responses to the Paradox of Credence	45
3.4.3 Responses to the Paradox of Knowledge	52
4 Non-Classical Solutions	54
4.1 Non-classical Solutions to the Paradox of Belief	54
4.1.1 The Structural Identity of Certain Doxastic and Semantic Paradoxes	54
4.1.2 The Paracomplete Solution	57
4.1.3 The Supervaluationist Solution	60
4.1.4 The Paraconsistent Solution	61
4.2 Non-classical Solutions to the Paradox of Credence	63
4.2.1 The Paracomplete Solution	63
4.2.2 The Supervaluationist Solution	68
4.2.3 The Paraconsistent Solution	72
4.3 Non-classical Solutions to the Paradox of Knowledge	74

4.3.1	The Paracomplete Solution	74
4.3.2	The Supervaluationist Solution	76
4.3.3	The Paraconsistent Solution	76
5	Non-Classical Attitudes	77
5.1	Non-Classical Doxastic and Credal States	77
5.1.1	Atomistic Contents and Non-Classical Doxastic/Credal States	79
5.1.2	Holistic Contents and Non-Classical Doxastic/Credal States	84
5.1.3	Rationality and Mental Content	85
5.1.4	Non-classical Credal States and Decision Theory	87
5.2	Non-Classical Epistemic States	89
6	Non-Classical Cognitive Significance	92
6.1	Paracomplete Cognitive Significance	93
6.1.1	Paracomplete Cognitive Significance: Qualitative Attitudes	93
6.1.2	Paracomplete Cognitive Significance: Quantitative Attitudes	111
6.2	Supervaluationist Cognitive Significance	118
6.3	Paraconsistent Cognitive Significance	120
	Bibliography	127

Acknowledgments

There are a number of people who deserve thanks in connection with this work.

My greatest intellectual debt from my time in graduate school is to John MacFarlane. Through instruction and example, John showed me how important philosophical work can come through careful philosophical engagement with formal theories.

Mike Martin also played an important role in helping to develop these and other related ideas. During his stints at Berkeley, Mike was a constant source of philosophical stimulation.

The main ideas for this work came while I was a visiting student at the University of Pittsburgh. Discussions with Cian Dorr at this time were extremely helpful for getting clear on some of the central theses of this work. Cian also deserves thanks for helpful discussions in the years since on this and related topics.

More recently, discussions with Hartry Field have been extremely useful in refining this material.

I was lucky to have a number of excellent fellow graduate students at Berkeley. Fabrizio Cariani and James Stazicker deserve special mention. They have both been constant sources of philosophical insight since my first year in grad school.

Thanks also go to Justin Bledin, Lara Buchak, John Campbell, Colin Caret, Stanley Chen, Kenny Easwaran, Melissa Fusco, Niko Koldony, Geoff Lee, Dilip Ninan, Alva Noe, Mike Rieppel, Dave Ripley, Sherri Roush, Paolo Santorio, Josh Sheptow, Shawn Standefer, Mike Titelbaum, Seth Yalcin, Brian Weatherson and Robbie Williams, and to audiences at the University of Toronto, the University of Rochester, Syracuse University, and the 2011 Formal Epistemology Workshop, for helpful conversations about, or comments on, the material in this and other related work.

I also gratefully acknowledge the support of the Social Science and Humanities Research Council of Canada.

Finally, my deepest thanks go to Jessica Gelber for being rad and awesome.

Chapter 1

Introductory Remarks

Consider the following sentence:

λ : $\ulcorner \lambda \urcorner$ is not true

As is well known, using a so-called liar sentence such as λ we can derive absurd results by appealing to classical logic, together with the inferences from ϕ to the claim that $\ulcorner \phi \urcorner$ is true and the converse. The principles of classical logic that we need to appeal to to derive these absurdities are very plausible. And so too are the semantic inferences involving truth. But what the liar paradox shows is that they aren't all correct. We must either revise our views about valid principles involving quantifiers and boolean connectives or we must revise our views about valid principles involving truth.

How we should revise our views in light of this incompatibility is a difficult and delicate matter. In this work, I don't try to weigh in on this question. Instead, the goal of this work is to explore certain ways in which questions about the nature of epistemic rationality and the mental states of rational agents, on the one hand, and questions about the correct treatment of the semantic paradoxes, on the other, can be mutually illuminating.

Certain answers to the question about how to correctly treat the semantic paradoxes can have surprising consequences for our understanding of doxastic rationality and the mental states of rational agents. In particular, treatments of the semantic paradoxes that abandon classical logic lead, I claim, to surprising views about the mental states of rational agents.

In addition to the semantic paradoxes, there are also paradoxes involving plausible normative principles governing doxastic and credal states. What these paradoxes show is that these principles, while *prima facie* very plausible, are classically inconsistent. Treatments of the semantic paradoxes that give up certain principles of classical logic offer sufficient resources to resolve these paradoxes. We can hold on to these seemingly fundamental principles, but doing so, however, requires allowing that the doxastic and credal states of rational agents will give rise to similar failures of classical logic as arise in cases of semantic paradox.

I claim that *if* one endorses a non-classical treatment of the semantic paradoxes, then one should also allow that these mental states will give rise to failures of classical logic. The argument

(very roughly) is as follows. We should accept the following (fairly weak) principle connecting rationality and doxastic/credal states: *ceteris paribus* we should prefer a description of an agent's mental state that maximizes rationality. Maximizing an agent's rationality requires, in certain cases at least, that the correct description of the agent's doxastic/credal state violate classical logic. Of course, the presumption in favor of the correctness of a rationalizing description would be outweighed if one were committed to the correctness of classical logic. However, if one allows for failures of classical logic, then I claim that the *ceteris paribus* preference for rationalizing characterizations should lead us to take such agents to have non-classical doxastic and credal states.

In this way, we can, I claim, potentially derive some surprising conclusions about the nature of the doxastic states of rational agents by appealing to seemingly rather distant considerations in the semantic realm.

Attention to certain questions about doxastic rationality is also important for understanding the significance of certain treatments of the semantic paradoxes. Here's an important question for any theory of semantic paradox: given that one accepts the theory, what attitude should one have towards the propositions expressed by paradoxical sentences such as $\ulcorner \lambda \urcorner$? Call this *the attitudinal question*. Part of our understanding of the significance of a theory of semantic paradox depends on how we think the attitudinal question should be answered.

In the case of non-classical treatments of the semantic paradoxes, I think that the answer to this question is not at all obvious; in fact, I think that the standard views about how proponents of various non-classical theories of semantic paradox should answer this question are mistaken.

The argument for this claim draws on the earlier discussion of our normative paradoxes. I show that although a proponent of a non-classical treatment of the semantic paradoxes has the resources to resolve these paradoxes, when we couple these theories with the standard answers to the attitudinal question, we can resurrect the normative paradoxes in a manner that is not amenable to similar solution. If, then, we want to hold on to some seemingly quite basic principles of doxastic/credal rationality, we must reject the standard answers about what attitude one should have towards paradoxical propositions given acceptance of a non-classical theory.

The answer to the attitudinal question, then, is constrained in a non-trivial way by certain basic principles of rationality. And what is particularly interesting, the answer is constrained in such a way that it rules out what have been seen as the most plausible answers to this question. In place of the standard answers to this question, I claim that if one endorses a non-classical treatment of the semantic paradoxes, then one should hold that rationality requires that one be such that the claim that one believes the proposition expressed by a paradoxical sentence has the same non-classical status as the paradoxical proposition itself.

That is some of what I'll be arguing for in this work. In more detail, here's what lies ahead.

In chapter 2, I very briefly lay out a number of different ways of responding to the liar paradox. The ways that I consider do not exhaust all the possible options. There are many theories that differ on subtle matters from the theories that I discuss. However, the theories that I consider serve as representatives of the most important classes of treatments of the semantic paradoxes. Moreover, those theories that differ from the one's that I consider will do so only in ways that should leave the main argumentative points I want to make untouched.

In chapter 3, I consider certain paradoxes that show that some very plausible normative principles governing doxastic and credal states are classically inconsistent. I also consider a (non-normative) paradox showing that some very plausible *descriptive* claims about epistemic states are classically inconsistent. I consider and criticize some standard responses to these paradoxes.

In chapter 4, I show how the non-classical theories developed in chapter 2 provide us with resources to solve these paradoxes.

In chapter 5, I take up the task of using the previously established results to argue for the view that doxastic, credal and epistemic states can give rise to failures of classical logic.

First, I consider whether doxastic/credal non-classicality is compatible with other commitments that we might have about the nature of such states. I argue that certain views about the nature of doxastic/credal states *may* be incompatible with doxastic/credal non-classicality. We can distinguish between *atomistic* and *holistic* theories of content. According to an atomistic theory, for any true ascription of the form ‘N.N. believes ϕ ’, there is a psychologically realized state with just ϕ as its content. According to a holistic theory, a doxastic state is simply something that carves out the space of possible worlds. An ascription of the form ‘N.N. believes ϕ ’, may be true in virtue of the space carved out by such a state, even if there is no psychologically realized bearer of just that content. I argue that (depending on how certain contentious questions are resolved) atomistic theories *may* be incompatible with doxastic and credal non-classicality. But, I show that there is no such incompatibility for holistic theories. Indeed, such non-classicality is shown to fall out of a very natural way of thinking about such states on the holistic view.

A minimal conclusion, then, is that there are plausible theories about the nature of doxastic and credal states that allow for doxastic non-classicality. I argue, however, that there are further attractive views about the nature of doxastic/credal states that, together with our earlier results, motivate the acceptance of doxastic/credal non-classicality. In particular, I argue that if one accepts certain plausible principles that accord rationality a constitutive role in determining the nature of doxastic and credal states, then we can marshal our earlier arguments to provide an argument that doxastic/credal states may provide an as yet unrecognized source of failures of classical logic. (Depending on how certain tricky issues are resolved, such constitutive principles may also, then, provide grounds for preferring a holistic theory of content to an atomistic theory.)

I argue, further, that the epistemic paradox developed in chapter 3 provides us with even clearer reason to allow for non-classical epistemic states. And I end the chapter by considering whether the role that doxastic/credal states play in the explanation of action is jeopardized by allowing for doxastic/credal non-classicality. The answer is no.

Finally, in chapter 6, I turn to the attitudinal question and discuss why it is important for our understanding the significance of theories of semantic paradox. I consider the standard answers to this question for the non-classical theories discussed in chapter 2. I then show how commitment to these standard answers serves to undermine the non-classical solutions to the epistemic paradoxes. Finally, I present alternative answers and show how these are compatible with the solutions to the normative paradoxes.

The picture that emerges from all of this is very different from the picture offered by standard non-classical treatments of the semantic paradoxes. Typically those who have been convinced that the semantic paradoxes show us that the principles of classical logic are not unrestrictedly valid

have thought that such non-classicality is restricted to the semantic domain. The lesson of this investigation is that this restriction is unmotivated. If there is non-classicality in the semantic domain, then there is non-classicality in the mental domain as well. Indeed, if there is non-classicality in the semantic domain, then this non-classicality will be perfectly mirrored in the doxastic and credal states of rational agents. The semantic paradoxes may, then, have more to teach us than has previously been thought.

Chapter 2

A Brief Introduction to Theories of Semantic Paradox

The liar paradox has a long history. In the last century, with the development of the tools of modern logic, treatments of this and related semantic paradoxes have reached a level of high mathematical sophistication. It would be well beyond the scope of this work to go through in detail even the main theories that have been developed over the last number of decades. The arguments that follow, however, will require at least some familiarity with certain technical treatments of the semantic paradoxes. Before, then, I proceed to the main argumentative material, it will be good to first briefly review the liar paradox, and to go over at least the fundamental elements of the various treatments of this paradox that will be relevant for the arguments that follow. This overview will be necessarily brief. However, for the interested reader I'll indicate, for each theory, other sources of greater illumination.

2.1 The Liar Paradox

A liar sentence is a sentence that asserts (in at least some sense) its own untruth. Let λ be a sentence such that:

$$(1) \models \lambda \leftrightarrow \neg T^{\ulcorner \lambda \urcorner}$$

Using this sentence we can derive a contradiction as follows:

¹ $\ulcorner \lambda \urcorner$ is a term that refers (perhaps under a certain coding scheme) to λ . It's worth noting the use I'll make of symbols such as ' T ', ' \neg ', ' \leftrightarrow ', ' $\ulcorner \lambda \urcorner$ ' etc. Throughout this work there will be two ways in which such symbols will be employed. On one use, such symbols are a part of a metalanguage which is used to talk about some object language. On this use ' T ', ' \neg ', ' \leftrightarrow ', ' $\ulcorner \lambda \urcorner$ ' etc. and concatenations of such symbols, will serve as *names* for expressions of an object language. On the second use, ' T ' is a predicate, ' \neg ' is a sentential connective, etc. whose interpretation should be obvious. Context should make it clear which of these two distinct uses is intended in any given case.

1	$\lambda \leftrightarrow \neg T^\Gamma \lambda^\neg$	(1)																				
2	$T^\Gamma \lambda^\neg \vee \neg T^\Gamma \lambda^\neg$	LEM																				
3	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;"> <table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">3 T-Elim</td> </tr> <tr> <td style="padding: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">1, 4 \leftrightarrow-Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">3, 5 \wedge-Intro</td> </tr> </table> </td> <td></td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;"> <table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">1, 7, \leftrightarrow Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">8, T-Intro</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">7, 9 \wedge-Intro</td> </tr> </table> </td> <td></td> </tr> </table>	<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">3 T-Elim</td> </tr> <tr> <td style="padding: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">1, 4 \leftrightarrow-Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">3, 5 \wedge-Intro</td> </tr> </table>	$T^\Gamma \lambda^\neg$		λ	3 T-Elim	$\neg T^\Gamma \lambda^\neg$	1, 4 \leftrightarrow -Elim	$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	3, 5 \wedge -Intro		<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">1, 7, \leftrightarrow Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">8, T-Intro</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">7, 9 \wedge-Intro</td> </tr> </table>	$\neg T^\Gamma \lambda^\neg$		λ	1, 7, \leftrightarrow Elim	$T^\Gamma \lambda^\neg$	8, T-Intro	$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	7, 9 \wedge -Intro		
<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">3 T-Elim</td> </tr> <tr> <td style="padding: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">1, 4 \leftrightarrow-Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">3, 5 \wedge-Intro</td> </tr> </table>	$T^\Gamma \lambda^\neg$		λ	3 T-Elim	$\neg T^\Gamma \lambda^\neg$	1, 4 \leftrightarrow -Elim	$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	3, 5 \wedge -Intro														
$T^\Gamma \lambda^\neg$																						
λ	3 T-Elim																					
$\neg T^\Gamma \lambda^\neg$	1, 4 \leftrightarrow -Elim																					
$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	3, 5 \wedge -Intro																					
<table style="border-collapse: collapse; width: 100%;"> <tr> <td style="border-bottom: 1px solid black; padding-bottom: 5px;">$\neg T^\Gamma \lambda^\neg$</td> <td></td> </tr> <tr> <td style="padding: 5px;">λ</td> <td style="padding-left: 10px;">1, 7, \leftrightarrow Elim</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">8, T-Intro</td> </tr> <tr> <td style="padding: 5px;">$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$</td> <td style="padding-left: 10px;">7, 9 \wedge-Intro</td> </tr> </table>	$\neg T^\Gamma \lambda^\neg$		λ	1, 7, \leftrightarrow Elim	$T^\Gamma \lambda^\neg$	8, T-Intro	$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	7, 9 \wedge -Intro														
$\neg T^\Gamma \lambda^\neg$																						
λ	1, 7, \leftrightarrow Elim																					
$T^\Gamma \lambda^\neg$	8, T-Intro																					
$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	7, 9 \wedge -Intro																					
11	$T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$	2, 3-6, 7-10 \vee -Elim																				

As we'll see, solutions to this paradox quickly become rather technical and complicated. The initial puzzle, however, is simple, and it leaves us with the following responses:

1. Reject the validity of T-Intro
2. Reject the validity of T-Elim
3. Reject the validity of the Law of Excluded-Middle (LEM)
4. Accept the conclusion
5. Reject the validity of \vee -Elim
6. Reject the validity of \leftrightarrow -Elim
7. Reject the validity of \wedge -Intro
8. Deny the existence of a sentence such as λ

Not all of these responses are on equal footing. 8, for example, is at least *prima facie* pretty unattractive. In natural language, there appear to be all sorts of devices by which sentences can achieve self-reference. For example, the complex demonstrative term *This sentence* would seem to provide us with a device by which we may construct sentences containing terms that refer to the very sentence of which the term is a constituent. Thus, in the sentence: *This sentence is a self-referential sentence*, the term *This sentence* would certainly appear to refer to that very sentence. We would seem, then, to be able to create a liar sentence using such complex demonstratives, e.g.,

*This sentence is not true.*² Moreover, the so-called diagonal lemma guarantees the existence of sentences such as λ satisfying the condition specified at (1), for languages capable of expressing fairly elementary arithmetic concepts.³

We won't, then, consider in detail any accounts that involve taking option 8. Nor we will consider any accounts that involve taking options 6 or 7. Perhaps there are ways of developing a general response to the semantic paradoxes that locate the problem in the above derivation either in the appeal to \leftrightarrow -Elim or in the appeal to \wedge -Intro. I'm not, however, aware of any such responses and so these options will not detain us.

For each of options 1 - 5, however, we'll see how we can develop a general theory of truth and logic that allows us to block the derivation of absurd conclusions. That such a *general* account is required should be obvious. For, of course, blocking one particular route to absurdity does not guarantee that there aren't other routes to be taken. For example, we can show that it isn't enough to merely reject excluded-middle. We can derive a contradiction without any appeal to this principle as follows:

1	$\lambda \leftrightarrow \neg T^{\ulcorner} \lambda^{\urcorner}$	(1)									
2	<table style="border-collapse: collapse; margin-left: 10px;"> <tr> <td style="border-right: 1px solid black; padding-right: 5px;"> </td> <td style="padding-left: 5px;">$T^{\ulcorner} \lambda^{\urcorner}$</td> <td> </td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;"> </td> <td style="padding-left: 5px;">λ</td> <td>2 T-Elim</td> </tr> <tr> <td style="border-right: 1px solid black; padding-right: 5px;"> </td> <td style="padding-left: 5px;">$\neg T^{\ulcorner} \lambda^{\urcorner}$</td> <td>1, 3 \leftrightarrow-Elim</td> </tr> </table>		$T^{\ulcorner} \lambda^{\urcorner}$			λ	2 T-Elim		$\neg T^{\ulcorner} \lambda^{\urcorner}$	1, 3 \leftrightarrow -Elim	
	$T^{\ulcorner} \lambda^{\urcorner}$										
	λ	2 T-Elim									
	$\neg T^{\ulcorner} \lambda^{\urcorner}$	1, 3 \leftrightarrow -Elim									
3	λ	2 T-Elim									
4	$\neg T^{\ulcorner} \lambda^{\urcorner}$	1, 3 \leftrightarrow -Elim									
5	$\neg T^{\ulcorner} \lambda^{\urcorner}$	2-4 \neg -Intro									
6	λ	1, 5 \leftrightarrow -Elim									
7	$T^{\ulcorner} \lambda^{\urcorner}$	6 T-Intro									
8	$T^{\ulcorner} \lambda^{\urcorner} \wedge \neg T^{\ulcorner} \lambda^{\urcorner}$	5, 7 \wedge -Intro									

What this shows is that certain theories that might be adequate to block the first derivation will nonetheless not provide an adequate general treatment of the liar paradox. For example, in an intuitionist logic not every instance of excluded-middle is valid. If we are intuitionists we need not accept the first derivation. However, intuitionists do take *reductio* reasoning to be a valid meta-rule, and so an intuitionist cannot block our second derivation.

A minimal demand, then, on an adequate treatment of the liar paradox is that it provide us with a *general* account of the validity relation for a language with a truth predicate and sufficient resources to achieve sentential self-reference that either (i) avoids leading to contradictions or (ii) if we are inclined to option 7, avoids leading from such contradictions to other even more manifestly absurd conclusions.

²Self-reference can also be achieved by means of definite descriptions, as well as by stipulation. For example it seems that I could just stipulate that *Jack* will denote the sentence *Jack is not true*. For discussion of this latter mode of self-reference see Kripke (1975).

³See Boolos and Jeffrey (1989).

Amongst options 1 - 5, a particularly salient division is between those options that require that we change our views about general logical principles and those that require that we change our views about general semantic principles. If we take either option 1 or 2, then we can hold on to classical logic. The cost, however, is that we need to reject certain intuitively valid inferences involving the notion of truth. If we take one of options 3 - 5, we can accept these inferences. However, in order to do this we need to reject intuitively plausible classically valid principles. In either case, we can't *both* hold on to classical logic and hold on to the validity of the inference from ϕ to $T\ulcorner\phi\urcorner$ and its converse. This much, at least, is a clear consequence of the liar paradox. What more we should say is a difficult question. Let us look at some of the possible options.

2.2 Treatments of the Liar Paradox

We'll consider a number of ways of responding to the liar paradox—each falling under one of options 1 - 5 above. We will proceed in two stages. First we will delineate certain classes of models. It is important to note that at this stage we're really only doing fairly simple mathematics. That is, there is nothing *contentious* happening at this stage. So, e.g., while in the discussion of our models we'll be talking about assignments to sentences of numbers, it is important that we don't at this point read any significance into these numbers. Having delineated various classes of models, however, we can then put the models to work in characterizing the validity relation for a first-order language. It is at this second stage that the difficult question arises: which if any of the characterizations accurately captures the validity relation for the language.⁴

2.2.1 Models

Classical Models

Let \mathcal{L} be a first-order language. \mathcal{L} contains variables: x_1, x_2, \dots and constants: c_1, c_2, \dots . For each sentence ϕ of \mathcal{L} we typographically single out a constant $\ulcorner\phi\urcorner$, which in the class of models that we will be considering will always be interpreted as denoting ϕ . Being a term of \mathcal{L} is defined in the standard way. Predicates of \mathcal{L} are: $P_1^1, P_2^1, \dots, P_1^2, P_2^2, \dots$, where superscript n indicates that the predicate is of arity n . Typographically we single out the unary predicate T . \mathcal{L} also has a first order quantifier: \forall , and boolean connectives: \wedge, \neg . The existential quantifier, disjunction and the material conditional are defined as follows: (i) $\exists x\phi =_{df} \neg\forall x\neg\phi$ (ii) $\phi \vee \psi =_{df} \neg(\neg\phi \wedge \neg\psi)$ (iii) $\phi \supset \psi =_{df} \neg\phi \vee \psi$. Finally, it will be useful later on to consider non-classical treatments for languages involving modal operators. We'll also assume, then, that \mathcal{L} has a sentential modal operator: \Box . Being a wff of \mathcal{L} and a sentence of \mathcal{L} are defined in the standard way.

Let \mathcal{L}^- be the language that results from \mathcal{L} by the removal of T . A classical model M^- for \mathcal{L}^- is a ordered tuple $\langle D_{m^-}, F, \Delta_{m^-}, R_{m^-}, \llbracket \rrbracket_{m^-} \rangle$. D_{m^-} is a domain of individuals. F is the set of assignment functions, i.e., the set of total functions from the set of variables of \mathcal{L}^- to D_{m^-} . Δ_{m^-}

⁴Here we're taking a model-theoretic approach to characterizing theories of semantic paradox. There is also an important axiomatic tradition. See, e.g., Tarski (1956), Friedman and Sheard (1987) and Halbach (2011).

is the set of points of evaluation. R_{m^-} is the accessibility relation on points in Δ_{m^-} . $\llbracket \cdot \rrbracket_{m^-}$ is an interpretation function. This maps triples of expressions (i.e., terms, predicates and wffs) of \mathcal{L}^- , assignment functions, and points of evaluation to semantic values. In a classical model M^- the interpretation function will satisfy the following constraints:

- For every constant c_i , every $g \in F$ and every $\delta \in \Delta_{m^-}$, $\llbracket c_i \rrbracket_{m^-}^{g,\delta} \in D_{m^-}$
- For every constant c_i , every $g, f \in F$, and every $\delta \in \Delta_{m^-}$, $\llbracket c_i \rrbracket_{m^-}^{g,\delta} = \llbracket c_i \rrbracket_{m^-}^{f,\delta}$
- For every predicate P_m^n , every $g \in F$, and every $\delta \in \Delta_{m^-}$, $\llbracket P_m^n \rrbracket_{m^-}^{g,\delta} \in F : D_{m^-}^n \mapsto \{0, 1\}$
- For every predicate P_m^n , every $g, f \in F$, and every $\delta \in \Delta_{m^-}$, $\llbracket P_m^n \rrbracket_{m^-}^{g,\delta} = \llbracket P_m^n \rrbracket_{m^-}^{f,\delta}$
- For every variable x_i , $\llbracket x_i \rrbracket_{m^-}^{g,\delta} = g(x_i)$
- For every wff ϕ , $g \in F$, and every $\delta \in \Delta_{m^-}$, if $\phi = P_i^n(t_1, \dots, t_n)$ then:

$$\llbracket \phi \rrbracket_{m^-}^{g,\delta} = \llbracket P_i^n \rrbracket_{m^-}^{g,\delta} < \llbracket t_1 \rrbracket_{m^-}^{g,\delta}, \dots, \llbracket t_n \rrbracket_{m^-}^{g,\delta} >$$
- For every wff ϕ , $g \in F$ and every $\delta \in \Delta_{m^-}$, if $\phi = \neg\psi$ then:

$$\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 1 \text{ iff } \llbracket \psi \rrbracket_{m^-}^{g,\delta} = 0$$

otherwise, $\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 0$
- For every wff ϕ , $g \in F$ and $\delta \in \Delta_{m^-}$, if $\phi = \psi \wedge \chi$ then:

$$\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 1 \text{ iff } \llbracket \psi \rrbracket_{m^-}^{g,\delta} = 1 \text{ and } \llbracket \chi \rrbracket_{m^-}^{g,\delta} = 1$$

otherwise, $\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 0$
- For every wff ϕ , $g \in F$, and $\delta \in \Delta_{m^-}$, if $\phi = \forall x_i \psi$ then:

$$\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 1 \text{ iff for every } g' \in F \text{ differing from } g \text{ at most with respect to } x_i, \llbracket \psi \rrbracket_{m^-}^{g',\delta} = 1$$

otherwise, $\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 0$
- For every wff ϕ , $g \in F$, and $\delta \in \Delta_{m^-}$, if $\phi = \Box\psi$ then:

$$\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 1 \text{ iff for every } \delta' \in \Delta_{m^-}, \text{ such that } \delta R_{m^-} \delta', \llbracket \psi \rrbracket_{m^-}^{g,\delta'} = 1$$

otherwise, $\llbracket \phi \rrbracket_{m^-}^{g,\delta} = 0$

In a classical model every formula will receive as a semantic value either 0 or 1, relative to any assignment function g and point of evaluation δ . We call the set $\{0, 1\}$ the *value space* of such a model.

A *sentence* is a formula with no free variables. Given a sentence ϕ , $\llbracket \phi \rrbracket_{m^-}^{g,\delta} = \llbracket \phi \rrbracket_{m^-}^{f,\delta}$, for all assignment functions g and f . We can then define the semantic value of a sentence ϕ in a model M^- , relative to a point of evaluation δ , $\llbracket \phi \rrbracket_{m^-}^\delta$, as the value ϕ receives relative to some (or equivalently: all) assignment function(s) at δ . More generally, we will employ the convention of

speaking of *the* semantic value of an expression at a point δ , when that expression is of a type that is not sensitive to the assignment function, e.g., constants and predicates. We'll denote this $\llbracket \cdot \rrbracket_{m^-}^\delta$.

In what follows we will always assume that M^- is a classical model respecting the following additional constraints:

- For every sentence ϕ of \mathcal{L} there is a term of \mathcal{L}^- , $\ulcorner \phi \urcorner$, such that for all $\delta \in \Delta_{m^-}$ $\llbracket \ulcorner \phi \urcorner \rrbracket_{m^-}^\delta = \phi$
(Note that in order to satisfy this stipulation D_{m^-} needs to contain all of the sentences of \mathcal{L} .)
- For every open formula with a single free variable $F(x)$ of \mathcal{L} there is a sentence $S = F \ulcorner S \urcorner$, i.e., a sentence which results from substituting for the free variable a term which, under M^- , refers to the sentence of which it is a part.
(This assumption ensures that \mathcal{L} has resources sufficient to achieve self-reference.)

Let us label the set of models for \mathcal{L}^- meeting all of the above constraints \mathcal{M}^- .

Non-Classical Models

A non-classical model is one with value space differing (in a non-trivial way) from $\{0, 1\}$. We will be interested in a number of different kinds of non-classical models here, but at first we will focus on non-classical models whose value space is $\{0, 1/2, 1\}$.

We'll say that such a non-classical model M for \mathcal{L} *extends* a classical model M^- for \mathcal{L}^- just in case the following conditions hold:

- (i) M and M^- have the same domain, the same set of points of evaluation, and the same accessibility relation.
- (ii) For every expression $\phi \in \mathcal{L}^-$, every $g \in F$, and $\delta \in \Delta_{m^-}$, $\llbracket \phi \rrbracket_m^{g,\delta} = \llbracket \phi \rrbracket_{m^-}^{g,\delta}$.

We will now outline two ways of extending a classical model M^- for \mathcal{L}^- to a non-classical model M for \mathcal{L} . The general method underlying these extensions is due to Saul Kripke.⁵ The key feature of the resultant non-classical models is that the semantic value assigned to ϕ is guaranteed to be the same as the value assigned to $T \ulcorner \phi \urcorner$, relative to each point of evaluation.

Non-Classical VF Models

To construct a non-classical model M^{vf} we consider a series of non-classical models M_α^{vf} . Before specifying how we construct such a series, we first provide a characterization of the class of models which will be potential members of the series.

⁵See Kripke (1975) for a more in depth treatment of this method of generating non-classical models. See also Field (2008), McGee (1991), and Soames (1999) for discussion of this construction. Note, however, that these authors do not consider models for languages involving modal operators. The differences that this addition occasions are, however, minimal.

- Each model M_α^{vf} is such that $D_{m_\alpha^{vf}} = D_{m^-}$, $\Delta_{m_\alpha^{vf}} = \Delta_{m^-}$, and $R_{m_\alpha^{vf}} = R_{m^-}$.
- For every $g \in F$, $\delta \in \Delta_{m_\alpha^{vf}}$ and constant c_i : $\llbracket c_i \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \llbracket c_i \rrbracket_{m^-}^{g,\delta}$
- For every predicate P_m^n of \mathcal{L} , every $g \in F$ and every $\delta \in \Delta_{m_\alpha^{vf}}$, $\llbracket P_m^n \rrbracket_{m_\alpha^{vf}}^{g,\delta} \in F : D_{m_\alpha^{vf}}^{g,\delta} \mapsto \{1, 1/2, 0\}$
- For every $g \in F$, and every $\delta \in \Delta_{m_\alpha^{vf}}$, for every predicate of \mathcal{L} other than T : $\llbracket P_m^n \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \llbracket P_m^n \rrbracket_{m^-}^{g,\delta}$
- For every $g, f \in F$, and every $\delta \in \Delta_{m_\alpha^{vf}}$, $\llbracket T \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \llbracket T \rrbracket_{m_\alpha^{vf}}^{f,\delta}$
- For every wff ϕ , $g \in F$, and every $\delta \in \Delta_{m_\alpha^{vf}}$, if $\phi = P_i^n(t_1, \dots, t_n)$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \llbracket P_i^n \rrbracket_{m_\alpha^{vf}}^{g,\delta} < \llbracket t_1 \rrbracket_{m_\alpha^{vf}}^{g,\delta}, \dots, \llbracket t_n \rrbracket_{m_\alpha^{vf}}^{g,\delta} >$$
- For every wff ϕ and $g \in F$, and $\delta \in \Delta_{m_\alpha^{vf}}$, if $\phi = \neg\psi$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{vf}}^{g,\delta} = 1 - \llbracket \psi \rrbracket_{m_\alpha^{vf}}^{g,\delta}$$
- For every wff ϕ , $g \in F$, and $\delta \in \Delta_{m_\alpha^{vf}}$, if $\phi = \psi \wedge \chi$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \min\{\llbracket \psi \rrbracket_{m_\alpha^{vf}}^{g,\delta}, \llbracket \chi \rrbracket_{m_\alpha^{vf}}^{g,\delta}\}$$
- For every wff ϕ , $g \in F$, and $\delta \in \Delta_{m_\alpha^{vf}}$, if $\phi = \forall x_i \psi$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \min\{\llbracket \psi \rrbracket_{m_\alpha^{vf}}^{g',\delta} : g' \in \Gamma_i\}, \text{ where } \Gamma_i \text{ is the set of assignment functions differing from } g \text{ at most with respect to } x_i.$$
- For every wff ϕ , $g \in F$, and $\delta \in \Delta_{m_\alpha^{vf}}$, if $\phi = \Box\psi$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{vf}}^{g,\delta} = \min\{\llbracket \psi \rrbracket_{m_\alpha^{vf}}^{s,\delta',M_\alpha^+} : \delta' \in \Delta_{m_\alpha^{vf}}^\delta\}, \text{ where } \Delta_{m_\alpha^{vf}}^\delta \text{ is the set of points } \delta' \text{ such that } \delta R_{m_\alpha^{vf}} \delta'.^6$$

It is easy enough to verify that these facts suffice to guarantee that M_α^{vf} will be an extension of M^- .

Given these stipulations, to construct our series of models we simply need to specify for each ordinal α the interpretation assigned to T by M_α^{vf} . We do so as follows.

- $\llbracket T \rrbracket_{m_0^{vf}}^\delta = f$: $f(x) = 0$, if x is not a sentence, and $f(x) = 1/2$ if x is a sentence.
- $\llbracket T \rrbracket_{m_{\beta+1}^{vf}}^\delta = f$: $f(x) = 0$ if either x is not a sentence, or x is a sentence ϕ and $\llbracket \phi \rrbracket_{m_\beta^{vf}}^\delta = 0$, $f(x) = 1$ if x is a sentence ϕ and $\llbracket \phi \rrbracket_{m_\beta^{vf}}^\delta = 1$, and otherwise $f(x) = 1/2$.

⁶Note that we assume that $\min(\emptyset) = 1$

- Where λ is a limit ordinal, $\llbracket T \rrbracket_{m_\lambda}^{\delta} = f$: $f(x) = 0$ if either x is not a sentence, or x is a sentence ϕ and there exists a $\gamma < \lambda$ such that $\llbracket \phi \rrbracket_{m_\gamma}^{\delta} = 0$, $f(x) = 1$ if x is a sentence ϕ and there exists a $\gamma < \lambda$ such that $\llbracket \phi \rrbracket_{m_\gamma}^{\delta} = 1$, and otherwise $f(x) = 1/2$.

This, then, serves to specify a (transfinite) sequence of non-classical models which extend M^- . I note the following important fact about this construction:

MONOTONICITY If $\gamma < \beta$ then $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0\} \subseteq \{x : \llbracket T \rrbracket_{m_\beta}^{\delta}(x) = 0\}$ and $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1\} \subseteq \{x : \llbracket T \rrbracket_{m_\beta}^{\delta}(x) = 1\}$.⁷

And the following is a simple consequence of MONOTONICITY:

FIXED POINT There is some γ such that for every $x \in D_{m_\gamma}^{\delta}$: $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1$ iff x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 1$; $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1/2$ iff x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 1/2$; and $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0$ iff either x is not a sentence or x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 0$.

Justification: To see that MONOTONICITY entails FIXED POINT note that according to MONOTONICITY as ordinal γ increases so does membership in $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0\}$ and $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1\}$. However, these sets cannot grow indefinitely. At most they could grow to contain all and only members of D_α^{δ} , since this set is the domain of all of the functions assigned to T by the models in our sequence. And since both Δ_α^{δ} and D_α^{δ} are sets there will be some ordinal γ of greater cardinality than the cardinality of $\Delta_\alpha^{\delta} \times D_\alpha^{\delta}$. By ordinal γ , $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0\}$ and $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1\}$ will have stopped growing. It follows that there will be an ordinal γ such that $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0\} = \{x : \llbracket T \rrbracket_{m_{\gamma+1}}^{\delta}(x) = 0\}$ and $\{x : \llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1\} = \{x : \llbracket T \rrbracket_{m_{\gamma+1}}^{\delta}(x) = 1\}$. But $\{x : \llbracket T \rrbracket_{m_{\gamma+1}}^{\delta}(x) = 0\}$ just is the set of x in D_{m_γ} such that either x is not a sentence or x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 0$, while $\{x : \llbracket T \rrbracket_{m_{\gamma+1}}^{\delta}(x) = 1\}$ just is the set of x such that x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 1$. It follows that there is an ordinal γ such that $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1$ iff x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 1$ and such that $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 0$ iff either x is not a sentence or x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 0$. And it follows from this that $\llbracket T \rrbracket_{m_\gamma}^{\delta}(x) = 1/2$ iff x is a sentence and $\llbracket x \rrbracket_{m_\gamma}^{\delta} = 1/2$. FIXED POINT thus follows from MONOTONICITY.

Let M_γ^{δ} be the first model in our construction with the properties specified in FIXED POINT.⁸ Here are a few relevant facts about our fixed-point model M_γ^{δ} :

⁷This can be proved by a simple inductive argument. See Kripke (1975).

⁸Since the ordinals are well-ordered there will be a first such model in our sequence.

- For every sentence ϕ of \mathcal{L}^- , $\llbracket \phi \rrbracket_{m^-}^\delta = \llbracket \phi \rrbracket_{m_y^{vf}}^\delta$
- For every sentence ϕ of \mathcal{L}^- , $\llbracket T \rrbracket_{m_y^{vf}}^\delta(\phi) = \llbracket \phi \rrbracket_{m^-}^\delta$
- And so for every sentence ϕ of \mathcal{L}^- , either $\llbracket T \rrbracket_{m_y^{vf}}^\delta(\phi) = 0$ or $\llbracket T \rrbracket_{m_y^{vf}}^\delta(\phi) = 1$
- For every sentence ϕ , $\llbracket \phi \rrbracket_{m_y^{vf}}^\delta = 1/2$ iff $\llbracket \neg\phi \rrbracket_{m_y^{vf}}^\delta = 1/2$
- There are sentences ϕ of \mathcal{L} such that $\llbracket \phi \rrbracket_{m_y^{vf}}^\delta = 1/2$

This last point requires some comment. Recall that we are assuming that for every formula of \mathcal{L} with a single free variable, $F(x)$, there is a sentence S such that $S = F^\top S^\top$. In particular, then, there is a sentence λ such that $\lambda = \neg T^\top \lambda^\top$. Given our stipulations about M^- , it follows that for every α $\llbracket \top \lambda^\top \rrbracket_{m_\alpha^{vf}}^\delta = \neg T^\top \lambda^\top$. It is easy enough to confirm that it follows from this that for every α $\llbracket T \rrbracket_{m_\alpha^{vf}}^\delta(\top \lambda^\top) = 1/2$ and so for every alpha $\llbracket T^\top \lambda^\top \rrbracket_{m_\alpha^{vf}}^\delta = \llbracket \neg T^\top \lambda^\top \rrbracket_{m_\alpha^{vf}}^\delta = 1/2$.

Recall our set of classical models \mathcal{M}^- . Let us label the set of models that one can get by applying the above construction to members of \mathcal{M}^- : \mathcal{M}^{vf} . In a moment we'll show how this class of models can be used in the treatment of the liar paradox. But first we'll look at a second way of constructing non-classical models for \mathcal{L} that extend members of \mathcal{M}^- .

Non-classical SV Models

To construct a non-classical model M^{sv} extending M^- , we again consider a series of non-classical models M_α^{sv} . Once again, we first provide a characterization of the class of models that will be potential members of the series and then show how to construct the series from members of this class.

Here we require a new notion. Let M be a model for \mathcal{L} with value space $\{1, 1/2, 0\}$. M , then, assigns to n-ary predicates of \mathcal{L} , relative to a point of evaluation, functions from D_m^n to $\{1, 1/2, 0\}$. Call M^\uparrow a *classical closure* of M just in case M^\uparrow is a classical model for \mathcal{L} such that (i) the domain of M^\uparrow is the same as the domain of M , $\Delta_{M^\uparrow} = \Delta_M$, and $R_{M^\uparrow} = R_M$, (ii) for every predicate P_m^n in \mathcal{L} and every $x \in D_m^n$ if $\llbracket P_m^n \rrbracket_m^\delta(x) = 1$ then $\llbracket P_m^n \rrbracket_{m^\uparrow}^\delta(x) = 1$, and if $\llbracket P_m^n \rrbracket_m^\delta(x) = 0$ then $\llbracket P_m^n \rrbracket_{m^\uparrow}^\delta(x) = 0$, and (iii) for every constant c_i in \mathcal{L} $\llbracket c_i \rrbracket_m^\delta = \llbracket c_i \rrbracket_{m^\uparrow}^\delta$. Given a non-classical model M , we label the set of classical closures of M : \mathcal{M}^\uparrow . We can now provide a characterization of the class of models that will be potential members, M_α^{sv} , in our series.

- Each model M_α^{sv} is such that $D_{m_\alpha^{sv}} = D_{m^-}$, $\Delta_{m_\alpha^{sv}} = \Delta_{m^-}$, and $R_{m_\alpha^{sv}} = R_{m^-}$.
- For every $g \in F$, $\delta \in \Delta_{m_\alpha^{sv}}$ and constant c_i : $\llbracket c_i \rrbracket_{m_\alpha^{sv}}^{g,\delta} = \llbracket c_i \rrbracket_{m^-}^{g,\delta}$
- For every predicate P_m^n of \mathcal{L} , $\delta \in \Delta_{m_\alpha^{sv}}$, and every $g \in F$, $\llbracket P_m^n \rrbracket_{m^-}^{g,\delta} \in F : D_{m^{vf}}^n \mapsto \{1, 1/2, 0\}$
- For every $g \in F$ and $\delta \in \Delta_{m_\alpha^{sv}}$, for every predicate of \mathcal{L} other than T : $\llbracket P_m^n \rrbracket_{m_\alpha^{sv}}^{g,\delta} = \llbracket P_m^n \rrbracket_{m^-}^{g,\delta}$

- For every $g, f \in F$, and $\delta \in \Delta_{m_\alpha^{sv}}$, $\llbracket T \rrbracket_{m_\alpha^{sv}}^{g,\delta} = \llbracket T \rrbracket_{m_\alpha^{sv}}^{g,\delta}$
- For every atomic wff ϕ , $g \in F$ and $\delta \in \Delta_{m_\alpha^{sv}}$, if $\phi = P_i^n(t_1, \dots, t_n)$ then:

$$\llbracket \phi \rrbracket_{m_\alpha^{sv}}^{g,\delta} = \llbracket P_i^n \rrbracket_{m_\alpha^{sv}}^{g,\delta} < \llbracket t_1 \rrbracket_{m_\alpha^{sv}}^{g,\delta}, \dots, \llbracket t_n \rrbracket_{m_\alpha^{sv}}^{g,\delta} >$$
- For every complex wff ϕ , $g \in F$ and $\delta \in \Delta_{m_\alpha^{sv}}$:

$$\llbracket \phi \rrbracket_{m_\alpha^{sv}}^{g,\delta} = 1 \text{ iff for every } M \in \mathcal{M}_{\alpha^{sv}}^\uparrow \llbracket \phi \rrbracket_M^{g,\delta} = 1$$

$$\llbracket \phi \rrbracket_{m_\alpha^{sv}}^{g,\delta} = 0 \text{ iff for every } M \in \mathcal{M}_{\alpha^{sv}}^\uparrow \llbracket \phi \rrbracket_M^{g,\delta} = 0$$
otherwise, $\llbracket \phi \rrbracket_{m_\alpha^{sv}}^{g,\delta} = 1/2$

It is easy enough to verify that these facts suffice to guarantee that M_α^{sv} will be an extension of M^- . Once again, to construct our series of models we need simply specify for each ordinal the interpretation of T . And this can be determined in exactly the same manner as before:

- $\llbracket T \rrbracket_{m_0^{sv}}^\delta = f$: $f(x) = 0$, if x is not a sentence, and $f(x) = 1/2$ if x is a sentence.
- $\llbracket T \rrbracket_{m_{\beta+1}^{sv}}^\delta = f$: $f(x) = 0$ if either x is not a sentence, or x is a sentence ϕ and $\llbracket \phi \rrbracket_{m_\beta^{sv}}^\delta = 0$, $f(x) = 1$ if x is a sentence ϕ and $\llbracket \phi \rrbracket_{m_\beta^{sv}}^\delta = 1$, and otherwise $f(x) = 1/2$.
- Where λ is a limit ordinal, $\llbracket T \rrbracket_{m_\lambda^{sv}}^\delta = f$: $f(x) = 0$ if either x is not a sentence, or x is a sentence ϕ and there exists a $\gamma < \lambda$ such that $\llbracket \phi \rrbracket_{m_\gamma^{sv}}^\delta = 0$, $f(x) = 1$ if x is a sentence ϕ and there exists a $\gamma < \lambda$ such that $\llbracket \phi \rrbracket_{m_\gamma^{sv}}^\delta = 1$, and otherwise $f(x) = 1/2$.

I note that, given this construction, MONOTONICITY holds, *mutatis mutandis*. And so FIXED POINT also holds, *mutatis mutandis*. Let M_γ^{sv} be the first model in our construction with the properties specified in FIXED POINT. The same facts that held for M_γ^{vf} also hold for our fixed-point model M_γ^{sv} :

- For every sentence ϕ of \mathcal{L}^- , $\llbracket \phi \rrbracket_{m^-}^\delta = \llbracket \phi \rrbracket_{m_\gamma^{sv}}^\delta$
- For every sentence ϕ of \mathcal{L}^- , $\llbracket T \rrbracket_{m_\gamma^{sv}}^\delta(\phi) = \llbracket \phi \rrbracket_{m^-}^\delta$
- And so for every sentence ϕ of \mathcal{L}^- , either $\llbracket T \rrbracket_{m_\gamma^{sv}}^\delta(\phi) = 0$ or $\llbracket T \rrbracket_{m_\gamma^{sv}}^\delta(\phi) = 1$
- For every sentence ϕ , $\llbracket \phi \rrbracket_{m_\gamma^{sv}}^\delta = 1/2$ iff $\llbracket \neg\phi \rrbracket_{m_\gamma^{sv}}^\delta = 1/2$
- There are sentences ϕ of \mathcal{L} , e.g., λ , such that $\llbracket \phi \rrbracket_{m_\gamma^{sv}}^\delta = 1/2$

What distinguishes M_γ^{sv} from M_γ^{vf} is how they treat logically complex sentences. Every classically valid formula will receive value 1 in M_γ^{sv} relative to a point δ , while every classical contradiction will receive value 0. This means that in certain cases a conjunction may have value 0 although both conjuncts have value 1/2. As an example, $T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$ will always have value 0, despite the fact that both conjuncts will have value 1/2. With M_γ^{vf} , however, a conjunction cannot have a lower value than the conjuncts, and so in this case the conjunction must have value 1/2 as well.

Let us label the set of models that one can get by applying the above construction to members of \mathcal{M} : \mathcal{M}^{sv} .

Classical Gap Models

Using \mathcal{M}^{vf} (or equally well \mathcal{M}^{sv}) we can derive a set of classical models that we'll call *classical gap models*.

Consider some member M_γ of \mathcal{M}^{vf} . We say that M^{ga} is a *gap closure* of M_γ just in case (i) M^{ga} is a classical model of \mathcal{L} that agrees with M_γ for all constants and predicates of \mathcal{L} other than T , and (ii) $\llbracket T \rrbracket_{m^{ga}}^\delta = f$: $f(x) = 1$ if $\llbracket T \rrbracket_{m_\gamma}^\delta(x) = 1$; otherwise $f(x) = 0$.

A sentence is a truth-value gap just in case it is neither true nor false, i.e., just in case neither it nor its negation is true. If a sentence has semantic value 1/2 in M_γ at δ then the claim that that sentence is a truth-value gap will have value 1 in M^{ga} at δ . To see this note that if $\llbracket \phi \rrbracket_{m_\gamma}^\delta = 1/2$ then $\llbracket \neg \phi \rrbracket_{m_\gamma}^\delta = 1/2$. And in this case $\llbracket T \rrbracket_{m^{ga}}^\delta(\phi) = \llbracket T \rrbracket_{m^{ga}}^\delta(\neg \phi) = 0$, and so $\llbracket \neg T^\Gamma \phi^\neg \rrbracket_{m^{ga}}^\delta = \llbracket \neg T^\Gamma \neg \phi^\neg \rrbracket_{m^{ga}}^\delta = \llbracket \neg T^\Gamma \phi^\neg \wedge \neg T^\Gamma \neg \phi^\neg \rrbracket_{m^{ga}}^\delta = 1$. In particular, then, the claim that liar sentence λ is a truth value gap has semantic value 1 according to M^{ga} .

Let \mathcal{M}^{ga} be the set of classical gap closures of members of \mathcal{M}^{vf} .

Classical Glut Models

Using \mathcal{M}^{vf} (or equally well \mathcal{M}^{sv}) we can also derive a set of classical models that we'll call *classical glut models*.

Consider some member M_γ of \mathcal{M}^{vf} . We say that M^{gl} is a *classical glut closure* of M_γ just in case (i) M^{gl} is a classical model of \mathcal{L} that agrees with M_γ for all constants and predicates of \mathcal{L} other than T , and (ii) $\llbracket T \rrbracket_{m^{gl}}^\delta = f$: $f(x) = 1$ if $\llbracket T \rrbracket_{m_\gamma}^\delta(x) = 1$ or $\llbracket T \rrbracket_{m_\gamma}^\delta(x) = 1/2$; otherwise $f(x) = 0$.

A sentence is a truth-value glut just in case it is both true and false, i.e., just in case both it and its negation are true. If a sentence has semantic value 1/2 in M_γ at δ then the claim that that sentence is a truth-value glut will have value 1 in M^{gl} at δ . To see this note that if $\llbracket \phi \rrbracket_{m_\gamma}^\delta = 1/2$ then $\llbracket \neg \phi \rrbracket_{m_\gamma}^\delta = 1/2$. And in this case $\llbracket T \rrbracket_{m^{gl}}^\delta(\phi) = \llbracket T \rrbracket_{m^{gl}}^\delta(\neg \phi) = 1$, and so $\llbracket T^\Gamma \phi^\neg \rrbracket_{m^{gl}}^\delta = \llbracket T^\Gamma \neg \phi^\neg \rrbracket_{m^{gl}}^\delta = \llbracket T^\Gamma \phi^\neg \wedge T^\Gamma \neg \phi^\neg \rrbracket_{m^{gl}}^\delta = 1$. In particular, then, the claim that the liar sentence λ is a truth value glut has semantic value 1 at δ according to M^{gl} .

Let \mathcal{M}^{gl} be the set of classical glut closures of members of \mathcal{M}^{vf} .

2.2.2 Validity

Classical Validity With Truth Value Gaps

Our first treatment of the liar paradox is one that holds on to classical logic and restricts certain plausible inferences involving the predicate T . Using \mathcal{M}^{ga} we can now provide the following putative characterization of the validity relation for \mathcal{L} . Let Γ be a set of sentences, and ϕ a sentence of \mathcal{L} . We say:

$\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{ga}$ and every $\delta \in \Delta_{m^{ga}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

Let us call the theory that endorses this as the correct account of validity for \mathcal{L} : KM .⁹ I note the following facts about KM :

- Since every model in \mathcal{M}^{ga} is a classical model, every classically valid inference is valid according to this account, and so too is every classical valid formula.
- The inference from $T^\Gamma \phi^\neg$ to ϕ is valid according to this account, i.e., we have $T^\Gamma \phi^\neg \models \phi$.

Justification: If M^{ga} is a classical gap closure of some $M_\gamma \in \mathcal{M}^{vf}$, we have $\llbracket T^\Gamma \phi^\neg \rrbracket_{m^{ga}}^\delta = 1$ iff $\llbracket \phi \rrbracket_{m_\gamma}^\delta = 1$. And if this latter holds, then we also have $\llbracket \phi \rrbracket_{m^{ga}}^\delta = 1$.

- The converse inference from ϕ to $T^\Gamma \phi^\neg$ is not valid according to this account, i.e., we have $\phi \not\models T^\Gamma \phi^\neg$.

Justification: For every $M_\gamma \in \mathcal{M}^{vf}$ $\llbracket \lambda \rrbracket_{m_\gamma}^\delta = \llbracket \neg T^\Gamma \lambda^\neg \rrbracket_{m_\gamma}^\delta = 1/2$. Thus for every $M^{ga} \in \mathcal{M}^{ga}$ $\llbracket \lambda \rrbracket_{m^{ga}}^\delta = \llbracket \neg T^\Gamma \lambda^\neg \rrbracket_{m^{ga}}^\delta = 1$, and $\llbracket T^\Gamma \lambda^\neg \rrbracket_{m^{ga}}^\delta = 0$.

- The claim that the liar sentence, λ , is a truth-value gap is valid, i.e., $\models \neg T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \neg \lambda^\neg$

We noted earlier that one way to block the derivation of a contradiction was to reject the validity of T-Intro, in particular the validity of the inference from λ to $T^\Gamma \lambda^\neg$. This was option 1. KM assures us that this is indeed sufficient to block the derivation of a contradiction in a language capable self-reference, i.e., that in giving up T-Intro we can hold on to all of the other intuitively plausible inference principles involved in the derivation of the contradiction. Moreover, KM assures us that in giving up the general validity of T-Intro we need not thereby forsake making ascriptions of truth. If we restrict our attention to the sentences of L^- , then the inference from ϕ to $T^\Gamma \phi^\neg$ will indeed count as valid according to this theory. Moreover, there will be plenty of sentences involving the truth predicate T that, according to KM , we can assert are true. There is, then, a fairly powerful restricted T-Intro principle that can be recovered from the theory. All of these are positive features of this approach to the liar paradox. We can hold on to classical logic

⁹I follow Field (2007) in the choice of label. See the final section of Kripke (1975) and Maudlin (2004) for a defense of this type of theory. See Field (2007) for critical discussion.

without being led into absurdity, while in addition being able to do justice to many intuitions about truth.

However, this theory does have certain unintuitive consequences. In certain cases, we cannot, according to KM, validly move from the assertion of a sentence to the claim that it is true. In fact, according to this theory in certain cases we must assert a sentence and also assert that that very sentence is *not* true, e.g., $\lambda \wedge \neg T^\Gamma \lambda^\neg$ is valid according to KM. In general, claiming both ϕ and that ϕ is not true sounds absurd. But, of course, since the liar paradox involves a set of claims all of which are intuitively plausible but which jointly lead to unacceptable conclusions any response to this paradox will involve accepting some unintuitive claims.

Classical Validity With Truth Value Gluts

Our next treatment of the liar paradox also holds on to classical logic, but does so by restricting the other direction of the T inferences.

Using \mathcal{M}^{gl} we can now provide the following putative characterization of the validity relation for \mathcal{L} . Let Γ be a set of sentences of \mathcal{L} and ϕ a sentence of \mathcal{L} . We say:

$\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{gl}$ and every $\delta \in \Delta_{m^{gl}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

Let us call the theory that endorses this as the correct account of the validity relation for \mathcal{L} : MK.¹⁰ I note the following facts about MK:

- Since every model in \mathcal{M}^{gl} is a classical model, every classically valid inference is valid according to this account, and so too is every classically valid formula.
- The inference from ϕ to $T^\Gamma \phi^\neg$ is valid according to this account, i.e., we have $\phi \models T^\Gamma \phi^\neg$.

Justification: If $\llbracket \phi \rrbracket_{m^{gl}}^\delta = 1$ then either $\llbracket \phi \rrbracket_{m^-}^\delta = 1$ or $\llbracket \phi \rrbracket_{m^-}^\delta = 1/2$. In either case we have $\llbracket T^\Gamma \phi^\neg \rrbracket_{m^{gl}}^\delta = 1$.

- The converse inference from $T^\Gamma \phi^\neg$ to ϕ is not valid according to this account, i.e., we have $T^\Gamma \phi^\neg \not\models \phi$.

Justification: For every $M_\gamma \in \mathcal{M}^{vf}$ $\llbracket \lambda \rrbracket_m^\delta = \llbracket \neg T^\Gamma \lambda^\neg \rrbracket_m^\delta = 1/2$. Thus for every $M^{gl} \in \mathcal{M}^{gl}$ $\llbracket \lambda \rrbracket_m^{gl\delta} = \llbracket \neg T^\Gamma \lambda^\neg \rrbracket_m^{gl\delta} = 0$, and $\llbracket T^\Gamma \lambda^\neg \rrbracket_m^{gl\delta} = 1$.

- The claim that the liar sentence, λ , is a truth-value glut is valid, i.e., $\models T^\Gamma \lambda^\neg \wedge T^\Gamma \neg \lambda^\neg$

We noted earlier that one way to block our derivation of a contradiction was to reject the validity of T-Elim, in particular the validity of the inference from $T^\Gamma \lambda^\neg$ to λ . This was option 2. MK assures us that this is indeed sufficient to block the derivation of a contradiction in a language

¹⁰See Field (2007) for critical discussion of this type of approach.

capable of self-reference, i.e., that in giving up T-Elim we can hold on to all of the other intuitively plausible inference principles involved in the derivation of the contradiction.

Unlike KM, if we accept MK we need not give up any of the intuitive claims that we want to make about which sentences are true. The problem for this theory is, of course, the converse; it commits us to claims about what is true that might seem perverse. In particular, there are cases, according to this theory, in which it is correct to claim that a particular sentence is true and yet also correct to deny that very sentence, i.e., to assert its negation. For example, $T^\Gamma \lambda^\neg \wedge \neg \lambda$ is valid according to MK.

Paracomplete Validity

In developing the above classical responses to the paradox, we employed the class of non-classical models \mathcal{M}^{vf} in order to determine the appropriate class of classical models, which we then used to provide a putative characterization of the validity relation for \mathcal{L} . Another way of responding to the paradoxes is to use \mathcal{M}^{vf} directly in the specification of the validity relation. Again, let Γ be a set of sentences and ϕ a sentence of \mathcal{L} . We say:

$\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{vf}$ and every $\delta \in \Delta_{m^{vf}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

Let us call the theory that endorses this as the correct account of the validity relation for \mathcal{L} : *KFS*.¹¹ I note the following facts about KFS:

- The validity relation according to KFS is non-classical.

For example excluded-middle fails to be a valid form on this account. In particular, we have $\not\models \lambda \vee \neg \lambda$. And more generally for any sentence ϕ that receives value 1/2 at some point δ in some $M \in \mathcal{M}^{vf}$, we have $\not\models \phi \vee \neg \phi$. The logic governing quantifiers and boolean connectives validated by KFS is called K_3 . This is just the logic induced by the so-called Strong Kleene valuation scheme, which is just the valuation that we used in determining the semantic values for our three-valued models in the Kripke construction, with semantic value 1 taken to be the sole *designated value*.

- We have $\phi \models T^\Gamma \phi^\neg$, i.e., the inference from ϕ to $T^\Gamma \phi^\neg$ is valid according to KFS.
Justification: For every $M \in \mathcal{M}^{vf}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket \phi \rrbracket_m^\delta = 1$ then $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta = 1$.
- We have $T^\Gamma \phi^\neg \models \phi$, i.e., the inference from $T^\Gamma \phi^\neg$ to ϕ is valid according to KFS.
Justification: For every $M \in \mathcal{M}^{vf}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta = 1$ then $\llbracket \phi \rrbracket_m^\delta = 1$.
- More generally let ψ and ϕ be sentences of \mathcal{L} and let ψ_ϕ be a sentence that results from substituting one or more occurrences of ϕ for $T^\Gamma \phi^\neg$ in ψ . Then we have, according to KFS, $\psi \models \psi_\phi$ and $\psi_\phi \models \psi$.

¹¹ See the early sections of Kripke (1975) for a development of this theory. See also Soames (1999), Richard (2008), and Field (2007) for endorsements of this type of theory.

Justification: It can be shown by a simple induction that for each $M \in \mathcal{M}^{vf}$ if ϕ and ψ are sentences that differ only in that one results from the other by the substitution of sentences which receive the same semantic value under M then $\llbracket \phi \rrbracket_m^\delta = \llbracket \psi \rrbracket_m^\delta$

KFS provides a general response to the liar paradox which gives up certain principles of classical logic and in return lets us hold on to the validity of T-Intro and T-Elim. We noted earlier that one way to block our derivation of a contradiction was to reject the validity of excluded-middle, in particular the validity of $T^\Gamma \lambda^\neg \vee \neg T^\Gamma \lambda^\neg$. This was option 3. KFS endorses this response. We also noted that simply rejecting excluded-middle is not in itself sufficient. There are ways of deriving a contradiction using other general principles endorsed by classical logic and the truth inference principles. In particular, we are able to avoid appeal to excluded-middle given appeal to *reductio* reasoning. In response to this alternative route to a contradiction, KFS rejects the validity of *reductio* reasoning. More precisely, the classical meta-rule: $\phi \models \neg\phi \Rightarrow \models \neg\phi$, is not sanctioned by KFS. To see this just substitute λ for ϕ .

Paraconsistent Validity

There is another non-classical way of responding to the paradoxes that uses the non-classical models in \mathcal{M}^{vf} to specify the validity relation. Instead of taking semantic value 1 to be the sole designated value, this account also takes semantic value 1/2 to be designated. Again, let Γ be a set of sentences and ϕ a sentence of \mathcal{L} . According to this account:

$\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{vf}$ and every $\delta \in \Delta_{m^{vf}}$, if $\llbracket \psi \rrbracket_m^\delta \geq 1/2$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta \geq 1/2$.

Let us call the theory that endorses this as the correct account of the validity relation for \mathcal{L} : LP.¹² I note the following facts about LP:

- Every *sentence* that is classically valid is valid according to LP.
- Certain sentences that are not classically valid are classically valid according to LP, e.g., $\models T^\Gamma \lambda^\neg \wedge \neg T^\Gamma \lambda^\neg$.
- However, not every *inference* that is classically valid is valid according to LP.

An important example here is the following. Classically we have: $\neg\phi, \phi \vee \psi \models \psi$. But according to LP this is not, in general, valid. As a counterexample let ϕ be λ and ψ be some contradiction in \mathcal{L}^- .

- We have $\phi \models T^\Gamma \phi^\neg$, i.e., the inference from ϕ to $T^\Gamma \phi^\neg$ is valid according to LP.

Justification: For every $M \in \mathcal{M}^{vf}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket \phi \rrbracket_m^\delta \geq 1$ then $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta \geq 1$.

¹²See Priest (2006b) for a defense and development of this type of account.

- We have $T^\Gamma \phi^\neg \models \phi$, i.e., the inference from $T^\Gamma \phi^\neg$ to ϕ is valid according LP.

Justification: For every $M \in \mathcal{M}^{vf}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta \geq 1$ then $\llbracket \phi \rrbracket_m^\delta \geq 1$.

- More generally let ψ and ϕ be sentences of \mathcal{L} and let ψ_ϕ be a sentence that results from substituting one or more occurrences of ϕ for $T^\Gamma \phi^\neg$ in ψ . Then we have, according to LP, $\psi \models \psi_\phi$ and $\psi_\phi \models \psi$.

Justification: It can be shown by a simple induction that for each $M \in \mathcal{M}^{vf}$ if ϕ and ψ are sentences that differ only one results from the other by the substitution of sentences which receive the same semantic value under M then $\llbracket \phi \rrbracket_m^\delta = \llbracket \psi \rrbracket_m^\delta$

We noted earlier that one response to our derivation of a contradiction was to simply accept the conclusion. This was option 4. A common response to this is to say that it is absurd to ever accept a contradiction. And, given classical logic, the absurdity is undeniable. For in classical logic *everything* follows from a contradiction. (To see this note that if we have a contradiction $\phi \wedge \neg\phi$ this gives us ϕ and $\neg\phi$. From the former we can infer $\phi \wedge \psi$, for any arbitrary ψ . But in classical logic we have $\neg\phi, \phi \vee \psi \models \psi$. And so in classical logic we can infer any arbitrary ψ from $\phi \wedge \neg\phi$.) According to the logic endorsed by LP, however, contradictions do not entail everything. And so if we adjust our views about logic in the way suggested by the proponent of LP, we need not agree that accepting a contradiction is absurd. And, in return for adjusting our views about logic in this way, we can, as with the case of KFS, maintain the validity of T-Intro and T-Elim.

Supervaluationist Validity

A fifth way of responding to the paradoxes is to use the non-classical models in \mathcal{M}^{sv} to specify the validity relation. Again, let Γ be a set of sentences of \mathcal{L} and ϕ a sentence of \mathcal{L} . According to this account:

$\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{sv}$ and every $\delta \in \Delta_{m^{sv}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

Let us call the theory that endorses this as the correct account of the validity relation for \mathcal{L} : SV.¹³ I note the following facts about SV:

- Every classically valid formula and every classically valid inference is valid according to SV.
- However, certain classically valid meta-rules are not valid according to this account.

Of note here are the following two classically valid meta-rules:

(\neg -Intro) $\phi \models \neg\phi \Rightarrow \models \neg\phi$

(\vee -Elim) $[(\phi \models \gamma) \wedge (\psi \models \gamma)] \Rightarrow \phi \vee \psi \models \gamma$

Both of these have counterexamples according to SV. Thus, according to SV: $\lambda \models \neg\lambda$, but $\not\models \neg\lambda$. While: $T^\Gamma \lambda^\neg \models \lambda$ and $\neg T^\Gamma \lambda^\neg \models \lambda$, but $T^\Gamma \lambda^\neg \vee \neg T^\Gamma \lambda^\neg \not\models \lambda$.

¹³See McGee (1991) for the development and defense of a similar theory. I also note that revision theories as developed e.g., in Gupta and Belnap (1993), share many of the features of this theory.

- We have $\phi \models T^\Gamma \phi^\neg$, i.e., the inference from ϕ to $T^\Gamma \phi^\neg$ is valid according to SV.
Justification: For every $M \in \mathcal{M}^{sv}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket \phi \rrbracket_m^\delta = 1$ then $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta = 1$.
- We have $T^\Gamma \phi^\neg \models \phi$, i.e., the inference from $T^\Gamma \phi^\neg$ to ϕ is valid according SV.
Justification: For every $M \in \mathcal{M}^{sv}$, $\llbracket \phi \rrbracket_m^\delta = \llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta$. So if $\llbracket T^\Gamma \phi^\neg \rrbracket_m^\delta = 1$ then $\llbracket \phi \rrbracket_m^\delta = 1$.

SV, like KFS and LP, provides a general response to the liar paradox that lets us hold on to the validity of T-Intro and T-Elim by giving up certain principles of classical logic. We noted earlier that we could block our first derivation of a contradiction if we gave up \vee -Elim. This was option 5. SV endorses this responses. And it can deal with the second derivation, since, as noted, it rejects the validity of *reductio* reasoning.

Although SV, KFS and LP require certain revisions in our views about logical validity, it may seem that SV is significantly more conservative in the revisions that it demands. After all, according to SV we need not revise any of our views about which sentences are valid, nor about which sentences are valid consequences of other sentences. It is only on matters of meta-rules, i.e., what claims of validity we can infer from other claims of validity, that SV demands revision.

It is worth making two points here.

The first point is that SV demands rejections of classically valid meta-rules that KFS and LP do not. SV, for example, rejects \neg -Elim, while LP does not. And SV also demands that we reject \vee – Elim, while this is accepted by both KFS and LP.

The second point to note, in this connection, is that although claims about meta-rules stated as claims about valid transitions between valid inferences may seem rather theoretical, such principles do correspond to certain modes of reasoning that are central to our understanding of logical validity. Meta-rules such as \neg -Elim and \vee – Elim correspond to natural modes of reasoning under supposition. Such modes of reasoning play a central role in our inferential lives. In weighing, then, the costs and benefits of SV, KFS and LP, we should not underestimate the cost associated with changing our views about which meta-rules are valid.¹⁴

2.2.3 Extensions

So far we've considered two classical, and three non-classical responses to the liar paradox. A problem with the non-classical theories as stated so far is that they are only defined for a language lacking certain expressive resources.

In \mathcal{L} , the only conditional-like connective is the material conditional: $\phi \supset \psi =_{df} \neg\phi \vee \psi =_{df} \neg(\phi \wedge \neg\psi)$. A problem with KFS and SV is that there are a number of conditional claims that are intuitively very plausible but that can't be expressed in those theories using \supset . In KFS we don't e.g., have $\models T^\Gamma \lambda^\neg \supset T^\Gamma \lambda^\neg$. And, in KFS and SV, although we have $T^\Gamma \lambda^\neg \models \lambda$, we don't have $\models T^\Gamma \lambda^\neg \supset \lambda$. Similarly, although in the these theories we have $\lambda \models T^\Gamma \lambda^\neg$, we don't have $\models \lambda \supset T^\Gamma \lambda^\neg$. Thus, although the T-inferences are valid according to KFS and SV, this fact cannot be encoded in a conditional within the language \mathcal{L} .

¹⁴See Field (2007) and Williamson (1994) for similar arguments.

A slightly different problem besets LP. In LP the above claims will hold. However, \supset fails to act as a suitable conditional for another reason, viz., *modus ponens* is not valid for this connective in LP. To see this, just recall that in order to allow for the truth of contradictions without being led into triviality, in LP we have: $\phi, \neg\phi \vee \psi \not\models \psi$. But this, of course, is just to say that in LP we have: $\phi, \phi \supset \psi \not\models \psi$, i.e., *modus ponens* is not valid in LP for \supset .

A further limitation of KFS and SV is the inability in \mathcal{L} to express the paradoxical status of e.g., λ . In LP, of course, we can say $\lambda \wedge \neg\lambda$. But in KFS and SV there is no sentence that would seem to characterize the unique status of λ that distinguishes it from non-paradoxical sentences, such as the sentences in \mathcal{L}^- . Of course, there is a lot that we can say in our *metalanguage*, but it would be nice if we could be assured that there was some way of expressing that λ is paradoxical in the object language in a manner that is consistent with the general principles of KFS or SV. And it will be useful for our purposes later on if we have such a device in the object language.

As it turns out, there is a way of remedying the shortcomings of these theories. To this end, I'll now outline a general method, due to Hartry Field, for generating models extending \mathcal{M}^{vf} for a language \mathcal{L}^+ that contains a conditional \rightarrow and a (defined) indeterminacy operator I that expresses the paradoxical status of the liar sentence. We can then use these models to provide putative characterizations of the validity relation for \mathcal{L}^+ extending KFS and LP. I'll then show how a similar method can be used to generate models extending \mathcal{M}^{sv} for \mathcal{L}^+ . These can then be used to provide a putative characterization of the validity relation for \mathcal{L}^+ extending SV.

Extended VF Models

We've seen how we can extend a classical model M^- for a language \mathcal{L}^- to a non-classical model M^{vf} for a language \mathcal{L} containing a predicate T . We'll now show how we can extend M^- to non-classical models M^{vf+} for a language \mathcal{L}^+ containing a predicate T as well as a conditional \rightarrow .

The construction, developed by Field, involves a transfinite sequence of Kripke constructions.¹⁵ At each stage in the sequence we begin with a model M^α . M^α assigns to the elements of language \mathcal{L}^- the assignments provided by M^- , it assigns to T the nullset as extension (relative each point δ) and, in addition, it assigns, relative to a sequences g and point δ , semantic values to formulas that have \rightarrow as their main connective. Given such a starting model we then construct a VF model M_+^α .

Consider an arbitrary formula with \rightarrow as its main connective: $\phi \rightarrow \psi$. The assignment $\llbracket \phi \rightarrow \psi \rrbracket_{M_+^\alpha}^{g,\delta}$ is determined as follows:

- For all g and δ , $\llbracket \phi \rightarrow \psi \rrbracket_{M^0}^{g,\delta} = 1/2$.
- For all g and δ , $\llbracket \phi \rightarrow \psi \rrbracket_{M_+^{\alpha+1}}^{g,\delta} = 1$ iff $\llbracket \phi \rrbracket_{M_+^\alpha}^{g,\delta} \leq \llbracket \psi \rrbracket_{M_+^\alpha}^{g,\delta}$; otherwise, $\llbracket \phi \rightarrow \psi \rrbracket_{M_+^{\alpha+1}}^{g,\delta} = 0$.
- For limit ordinal λ , for all g and δ , $\llbracket \phi \rightarrow \psi \rrbracket_{M_+^\lambda}^{g,\delta} = 1$ iff there exists some stage $\beta < \lambda$ such that for all $\sigma, \beta \leq \sigma < \lambda$, $\llbracket \phi \rrbracket_{M_+^\sigma}^{g,\delta} \leq \llbracket \psi \rrbracket_{M_+^\sigma}^{g,\delta}$.

¹⁵See Field (2008), Field (2007) for a more in depth account of this method of construction.

- For limit ordinal λ , for all g and δ , $\llbracket \phi \rightarrow \psi \rrbracket_{M^\lambda}^{g,\delta} = 0$ iff there exists some stage $\beta < \lambda$ such that for all $\sigma, \beta \leq \sigma < \lambda$, $\llbracket \phi \rrbracket_{M_\sigma^+}^{g,\delta} > \llbracket \psi \rrbracket_{M_\sigma^+}^{g,\delta}$.
- Otherwise, $\llbracket \phi \rightarrow \psi \rrbracket_{M^\lambda}^{g,\delta} = 1/2$

At each stage α , then, a formula ϕ will receive a semantic value relative to a sequence g and point δ , given the resulting VF model at that stage M_α^+ . An important point to note is that the assignments to formulas of values relative to $\langle g, \delta \rangle$ pairs will eventually fall into a cyclical pattern. This is a consequence of the fact that there are ordinals of greater cardinality than the cardinality of the set of possible functions from formula, sequence, point triples to values 1, 0, 1/2. At some point, then, a valuation will recur and this will institute a cyclical pattern.

Instead of taking the semantic value of a formula ϕ (relative to a sequence g and point δ) to be either 1, 1/2, or 0 (as in the case of Kripke models) we take it to be the transfinite sequence of values 1, 1/2, 0 that the formula eventually cycles on. More specifically, if ξ is the ordinal length of the valuation cycle given our construction, then we'll let $\llbracket \phi \rrbracket_{M^+}^{g,\delta} = f$, where f is a member of $F : \{\beta : \beta \in \xi\} \mapsto \{1, 1/2, 0\}$, and f maps each $\beta \in \xi$ to the value that ϕ receives relative to g and δ at that point in the cycle.

The value space of our model, then, is a subset of $F : \{\beta : \beta \in \xi\} \mapsto \{1, 1/2, 0\}$, i.e., the set of functions from ordinals in ξ to values 1, 1/2, 0. We can define the following partial-ordering on this set of values. Where f and g are members of $F : \{\beta : \beta \in \xi\} \mapsto \{1, 1/2, 0\}$, we say that $f \leq g$ iff for all $\beta \in \xi$ $f(\beta) \leq g(\beta)$. Within this value space there will be a top-value viz., f : for all $\beta \in \xi$ $f(\beta) = 1$. There will also be a bottom value, viz., f : for all $\beta \in \xi$ $f(\beta) = 0$. In addition there will be a ‘‘middle-value’’ viz., f : for all $\beta \in \xi$ $f(\beta) = 1/2$. This latter behaves like value 1/2 in VF models in that whenever a formula has this value so does its negation. Where there is no possibility of confusion I will refer to these values as: 1, 1/2, 0.

Let \mathcal{M}^{vf^+} be the class of models that results from constructing VF models at each stage of this sequence. I note a few important facts about this class of models:

- If ϕ is a formula of L^- , then $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = 1$ or $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = 0$.
- If ϕ is a formula of \mathcal{L} , i.e., a formula not containing \rightarrow , then $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = 1$, or $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = 1/2$ or $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = 0$.
- For every formula ϕ , $\llbracket \phi \rrbracket_{M^{vf^+}}^{g,\delta} = \llbracket T^\Gamma \phi^\neg \rrbracket_{M^{vf^+}}^{g,\delta}$.

Validity

Using the class of models \mathcal{M}^{vf^+} , we can now provide the following putative characterization of the validity relation for L^+ .

KFS⁺ $\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{vf^+}$ and every $\delta \in \Delta_{m^{vf^+}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

I note the following facts about KFS^+ :

- As with KFS, certain classically valid formulas are not valid according to KFS^+ . For example, excluded-middle fails to be valid, i.e., $\not\models \phi \vee \neg\phi$.
- We also get failures of classically valid meta-rules such as *reductio*.
- As with KFS, the T-Inferences are valid. We have $\phi \models T^\Gamma \phi^\neg$ and $T^\Gamma \phi^\neg \models \phi$.
- Unlike with KFS, however, KFS^+ is defined for a language in which facts about the equivalence between ϕ and $T^\Gamma \phi^\neg$ can be encoded. In particular we have: $\models T^\Gamma \phi^\neg \leftrightarrow \phi$.
- We also have $\models I\lambda$. That is, the claim that the liar sentence is indeterminate is valid according to this theory. In \mathcal{L}^+ we can express this status of the liar sentence using an indeterminacy operator.¹⁶

LP⁺ $\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{vf^+}$ and every $\delta \in \Delta_{m^{vf^+}}$, if $\llbracket \psi \rrbracket_m^\delta \geq 1/2$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta \geq 1/2$.

I note the following facts about LP^+ :

- As with LP, certain contradictions are, according to LP^+ valid, e.g., we have $\models \lambda \wedge \neg\lambda$.
- As with LP, the T-inferences are valid. We have $\phi \models T^\Gamma \phi^\neg$ and $T^\Gamma \phi^\neg \models \phi$.
- Unlike with LP, the conditional available in \mathcal{L}^+ is one for which modus ponens holds, i.e., we have $\phi, \phi \rightarrow \psi \models \psi$.

KFS^+ and LP^+ , like KFS and LP, represent approaches to the semantic paradoxes that recommend tackling the liar paradox in manners 3 and 4 respectively. What makes the former theories preferable to the latter is simply that they are defined for a language with greater expressive resources. These resources allow us to reason about paradoxical cases in ways that simply aren't available in the languages for which KFS and LP are defined.

Extended SV models

It might seem that we can just as easily apply this method of construction to create models extending SV models. Instead of generating our models using transfinite sequences of VF models, why not generate models using transfinite sequences of SV models? In fact, this particular method of generating models for \mathcal{L}^+ won't provide us with models that can be used to characterize a reasonable extension of SV. The reason for this is that formulas with \rightarrow are treated as atomic formulas at each stage in our transfinite construction. They are simply assigned a semantic value and that value doesn't depend on the values assigned to any constants or predicates in the models at

¹⁶Indeterminacy can be defined as follows in the language: $I\phi =_{df} \neg(\phi \wedge \neg(\phi \rightarrow \neg\phi)) \wedge \neg(\neg\phi \wedge \neg(\neg\phi \rightarrow \neg\neg\phi))$.

that stage in the construction. What this means is that if a formula of the form $\phi \rightarrow \psi$ has semantic value $1/2$ in the base model at some stage in our construction then it will have $1/2$ also relative to every “classical closure”, since its value does not depend on the extension of any predicates. And so $(\phi \rightarrow \psi) \vee \neg(\phi \rightarrow \psi)$, won’t receive value 1 in every classical closure.

The way to resolve this problem is to redefine what we mean by *classical closure*. In discussing \mathcal{L} , we said that given a model M with value-space $\{1, 1/2, 0\}$, M^\uparrow was a classical closure of M just in case (i) the domain of M^\uparrow is the same as the domain of M , $\Delta_{M^\uparrow} = \Delta_M$, and $R_{M^\uparrow} = R_M$, (ii) for every predicate P_m^n in \mathcal{L} and every $x \in D_m^n$ if $\llbracket P_m^n \rrbracket_m^\delta(x) = 1$ then $\llbracket P_m^n \rrbracket_{m^\uparrow}^\delta(x) = 1$, and if $\llbracket P_m^n \rrbracket_m^\delta(x) = 0$ then $\llbracket P_m^n \rrbracket_{m^\uparrow}^\delta(x) = 0$, and (iii) for every constant c_i in \mathcal{L} $\llbracket c_i \rrbracket_m^\delta = \llbracket c_i \rrbracket_{m^\uparrow}^\delta$. Now that we are dealing with models for \mathcal{L}^+ , we need to add an additional clause. Let Π be the set of formulas with \rightarrow as their main connective. Then, given a model M for \mathcal{L}^+ with value-space $\{1, 1/2, 0\}$, we say that M^\uparrow is a classical closure of M just in case clauses (i) - (iii) hold, and in addition for every $\phi \in \Pi$ (iv) either $\llbracket \phi \rrbracket_{m^\uparrow}^\delta = 1$ or $\llbracket \phi \rrbracket_m^\delta = 0$ (but not both), (v) if $\llbracket \phi \rrbracket_m^\delta = 1$, then $\llbracket \phi \rrbracket_{m^\uparrow}^\delta = 1$, (vi) if $\llbracket \phi \rrbracket_m^\delta = 0$, then $\llbracket \phi \rrbracket_{m^\uparrow}^\delta = 0$.

Given an assignment of objects to constants, functions from n-tuples to the value space $\{0, 1/2, 1\}$ to predicates P_m^n , and values from the space $\{0, 1/2, 1\}$ to members of Π , we can then characterize an SV model for \mathcal{L}^+ using our new notion of a classical closure. Given this new set of SV models we can construct Kripke models using the fixed-point construction in the same manner as we did before. Finally, given some starting SV Kripke model, we can construct a transfinite revision sequence in the manner specified above. This will differ from the VF case only in that at each stage of the construction our base assignments will be used to construct an SV model instead of a VF model.

Let us label the class of models that result from applying this transfinite construction to SV models: \mathcal{M}^{sv^+} .

Here are a few relevant facts about these models:

- If ϕ is a formula of L^- , then $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = 1$ or $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = 0$.
- If ϕ is a formula of \mathcal{L} , i.e., a formula not containing \rightarrow , then $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = 1$, or $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = 1/2$ or $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = 0$.
- For every formula ϕ , $\llbracket \phi \rrbracket_{M^{sv^+}}^{g,\delta} = \llbracket T^\Gamma \phi^\neg \rrbracket_{M^{sv^+}}^{g,\delta}$.

Validity

Using the class of models \mathcal{M}^{sv^+} , we can now provide the following putative characterization of the validity relation for L^+ .

SV⁺ $\Gamma \models \phi$ iff for every $M \in \mathcal{M}^{sv^+}$ and every $\delta \in \Delta_{m^{sv^+}}$, if $\llbracket \psi \rrbracket_m^\delta = 1$ for every $\psi \in \Gamma$, then $\llbracket \phi \rrbracket_m^\delta = 1$.

I note the following facts about SV⁺:

- As with SV, every classically valid formula is valid according to SV^+ .
- As with SV, we get failures of classically valid meta-rules, such as proof-by-cases.
- As with SV, the T-inferences are valid. We have $\phi \models T^\Gamma \phi^\neg$ and $T^\Gamma \phi^\neg \models \phi$.
- $I\lambda$ is valid according to this theory, and so, once again, we can express the paradoxical status of the liar sentence using the language \mathcal{L}^+ .

2.3 Going Ahead

A paradox is an argument that shows that (at least) one of a set of *prima facie* plausible claims cannot be true, but without providing any guidance as to which of the claims is false. The derivation of a contradiction from classical logic and the T-inferences is a paradigmatic case of a paradox. Here the set of claims that are jointly unacceptable are all individually *extremely* plausible. And yet the argument that shows that at least one of these assumptions is false doesn't in itself provide us with any clue as to which of the assumptions is ultimately unacceptable.

So far I've canvassed five types of response to the liar paradox, and shown how they can be developed into rigorous accounts of the validity relation for a language involving a truth predicate. Two of these responses involved giving up at least one of the T-inferences. In exchange, such responses allow us to hold on to classical logic. The other three types of response involve giving up certain elements of classical logic. In exchange, such responses allow us to hold on to the validity of the inferences from ϕ to $T^\Gamma \phi^\neg$ and its converse.

It is a *very* difficult question which of these is the correct way of responding to the paradox. In our present state of knowledge, it seems to me that we should give *roughly* equal credence to each of these ways of responding to the paradox. I say *roughly* here, because I think that when we compare particular theories advocating these types responses we can see that there may be some reasons for preferring one theory to another. But the sorts of considerations available for the most part turn on rather subtle questions about how we should weigh various costs and benefits of the respective theories. In response to such data, it is, I think, most reasonable to (roughly) divide one's credences. I don't think, then, that the liar paradox clearly shows us that we must revise our beliefs about what general principles govern the boolean operators and quantifiers. I do, however, think that this paradox makes certain alternatives to classical logic reasonable epistemic possibilities.

In what follows, I will be arguing for various *conditional* claims. I'll be arguing for claims of the form: *if* certain non-classical theories are correct, then certain consequences follow for our understanding of mental states and rationality. Such conditional claims are interesting because their antecedents might very well be true and their consequents are rather surprising.

The non-classical theories that I'll focus on will be KFS^+ , LP^+ and SV^+ , with particular attention being paid to the first of these. It's important that the arguments be targeted at particular theories. Given that we're considering various revisions to our pretheoretic views about which inferences are valid, it is important that we be clear what inferential moves are ok to appeal to in developing our arguments. KFS^+ , LP^+ and SV^+ provide clear constraints on the inferences that we

can make. Other accounts may agree with, say, KFS^+ in rejecting excluded-middle and reductio as valid principles, but nonetheless disagree with KFS^+ with respect to other principles.¹⁷ So the fact that certain claims hold for these theories does not mean that they automatically hold for any theory that endorses a revision to classical logic in response to the liar paradox. Nonetheless, the inferential moves that are required for the conclusions that I want to establish are fairly minimal. If, however, one endorses an alternative non-classical account, one will need to check that the arguments that follow can indeed be reframed within that theory. In most cases, this should be possible. But it would be futile to try to argue for this in general. The space of conceivable logical alternatives is simply too vast.

¹⁷See e.g., Yablo (2003), for an account which disagrees with KFS^+ over the logic governing \rightarrow .

Chapter 3

Attitudinal Paradoxes

3.1 Belief

Consider the following sentence:

I do not believe that this sentence is true¹

What should your attitude be towards this sentence? You're likely to be puzzled. You know the following facts:

- If you believe that it's true, then it's false.
- If you don't believe that it's true, then it's true.

Assuming that you'll know whether or not you believe that it's true, you'll then either be in the position of knowing that you believe that the sentence is true, and knowing that your so believing makes it false, or knowing that you fail to believe that the sentence is true, and knowing that your so failing to believe makes it true. Neither of these seems like a rational state for an agent to be in.

Our puzzlement at this case can be sharpened into a paradox. In a moment, I'll show how this works in precise detail, but let me first give you a sense of the form that the paradox takes.

Using the type of sentence above, we can argue that there is a possible agent who, without being guilty of any antecedent rational failing, is unable to satisfy the following two plausible normative principles:

CONSISTENCY For any proposition ϕ , it is a rational requirement that if one believes ϕ then one not believe its negation $\neg\phi$.²

¹This type of sentence and some of its odd features are discussed in Burge (1978), Burge (1984), Conee (1987) and Sorensen (1988).

²Read: $O(B\phi \rightarrow \neg B\neg\phi)$.

EVIDENCE For any proposition ϕ , if an agent's evidence makes ϕ certain then the agent is rationally required to believe ϕ .³

If our agent believes that the above sentence is true, then it will either fail to satisfy CONSISTENCY or EVIDENCE. And, if our agent doesn't believe that the above sentence is true, then it will fail to satisfy EVIDENCE. In either case, our agent will be guilty of a rational failure. What this would seem to show is that CONSISTENCY and EVIDENCE are incompatible with the following general constraint on principles of rationality:

POSSIBILITY It must always be possible for an antecedently rational agent to continue to meet the requirements imposed on it by rationality.

Let us see how this paradox works in detail.

Let B_α be an operator meaning *Alpha believes that*. Let $\ulcorner\beta\urcorner$ name the following sentence: $\neg B_\alpha T\ulcorner\beta\urcorner$.⁴ Then, as an instance of the T-schema, we have:

$$(2) T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)^5$$

Actual agents are good at detecting their own doxastic states. This works in two directions. First, when one believes something, one often believes that one believes it; similarly when one does not believe something, one often believes that one does not believe it. Second, when one believes that one believes something, for the most part one is right; similarly when one believes that one does not believe something. Still, actual agents are fallible in both directions. There are plenty of beliefs of mine of which I am unaware, and which would remain hidden to me even after a thorough introspective search, and the same is, I take it, true of you. There are also beliefs that I have about my own belief state that are false. While it may have once been common to suppose that each agent's mind is transparent to herself, this thought now seems indefensible.

Many of our limitations in this respect would, however, seem to be medical in nature, not metaphysical. Consider an ideal agent. Call it *Agent Alpha*. The following seems, at the very least, metaphysically possible. Whenever Alpha believes that $\ulcorner\beta\urcorner$ is true, then it also believes that it believes it. And, whenever Alpha does not believe that $\ulcorner\beta\urcorner$ is true, then it believes that it does not believe it. Moreover, Alpha is, overall, *perfectly* reliable in the higher-order beliefs that it has with respect to whether or not it believes that $\ulcorner\beta\urcorner$ is true. Alpha believes that it believes that $\ulcorner\beta\urcorner$

³Two points. (i) Note that EVIDENCE is a *synchronic* norm. If, at a particular time t an agent has evidence that makes ϕ certain and fails to believe it, then the agent is thereby subject to rational criticism. (ii) I take it that there are weaker levels of evidential support that also rationally mandate belief. It is, however, neater to work with this (logically) weaker rational constraint. But note that if one is uncomfortable with the idea that one's evidence ever makes anything certain, the following puzzle could be recreated by appeal to a plausible normative constraint to the effect that there is some less than conclusive evidential threshold beyond which belief is rationally mandated.

⁴Here, following Kripke (1975), sentential self-reference is achieved by stipulation. This could also, of course, be achieved by a technique such as Gödel numbering.

⁵Note that \leftrightarrow should here just be read as an intuitive conditional. We'll worry later about what valid principles might govern such a conditional and whether the particular claims that we're making here can ultimately be sustained given a precise interpretation of \rightarrow .

is true only if it does believe that $\ulcorner\beta\urcorner$ is true, and it believes that it does not believe that $\ulcorner\beta\urcorner$ is true only if it does not believe that $\ulcorner\beta\urcorner$ is true.

More perspicuously, then, we have the following:

$$(3) B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow B_\alpha B_\alpha T(\ulcorner\beta\urcorner)$$

$$(4) \neg B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)^6$$

We may further suppose that our agent believes the truth expressed in (2). We have then:

$$(5) B_\alpha (B_\alpha T(\ulcorner\beta\urcorner) \rightarrow \neg T(\ulcorner\beta\urcorner))$$

$$(6) B_\alpha (\neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow T(\ulcorner\beta\urcorner))$$

The possibility of an agent, such as Alpha, who satisfies (3)-(6), raises problems for the conjunction of CONSISTENCY, EVIDENCE and POSSIBILITY. To see this, first consider the following two cases.

Case 1: On the assumption that Alpha does not believe that β is true, it follows that it ought to believe that β is true.

Assume that Alpha does not believe that $\ulcorner\beta\urcorner$ is true. By (4), it follows that it believes that it does not believe this. Alpha also believes that if it does not believe this then $\ulcorner\beta\urcorner$ is true. This is (6). The set-up of the case is such that the status of both of these beliefs is superlative. The first belief is *perfectly* reliable. The second proposition that it believes is a theorem, and we can assume that its grounds for believing this are the same as ours. Given their *bona fides*, these beliefs, I claim, form part of the agent's total body of evidence.⁷ We can assume, further, that Alpha has no other evidence that bears one way or the other on the question of whether $\ulcorner\beta\urcorner$ is true. Alpha, then, would seem to be in the position in which its evidence makes it certain that $\ulcorner\beta\urcorner$ is true. By EVIDENCE, it follows that Alpha ought to believe that $\ulcorner\beta\urcorner$ is true.

⁶We should think of B_α as involving a first-personal mode of presentation for Alpha.

⁷Exactly what sort of relation one must bear to a proposition in order for the latter to be part of one's evidence is a topic of some controversy. The case, however, is set-up so that Alpha should meet any reasonable standards. Alpha, for example, knows that it does not believe that $\ulcorner\beta\urcorner$ is true, and that if it does not believe that $\ulcorner\beta\urcorner$ is true then $\ulcorner\beta\urcorner$ is true. Our assumption, then, is justified if one holds that a proposition ϕ counts as part of an agent's evidence just in case the agent knows that ϕ . See, for example, Williamson (2000). In addition, the agent's knowledge in both cases need not be inferential. Our assumption, then, is justified if one thinks that that a proposition ϕ counts as part of an agent's evidence just in case the agent knows that ϕ and ϕ is not inferred from other known premisses. And, of course, our assumption is justified *a fortiori* if one thinks that a proposition ϕ that one believes counts as evidence just in case one satisfies some less demanding criteria, for example, having a justified belief in ϕ . See, for example, Feldman (2004).

Case 2: On the assumption that Alpha does believe that $\ulcorner\beta\urcorner$ is true, it follows that it ought not believe that $\ulcorner\beta\urcorner$ is true.

Assume that Alpha does believe that $\ulcorner\beta\urcorner$ is true. By (3), it follows that it believes that it believes that $\ulcorner\beta\urcorner$ is true. Alpha also believes that if it believes that $\ulcorner\beta\urcorner$ is true then $\ulcorner\beta\urcorner$ is not true. This is (5). Again the evidential status of these beliefs is, by the set-up of the case, superlative. We can again assume that Alpha has no other evidence that bears on whether or not $\ulcorner\beta\urcorner$ is true. Alpha, then, is such that its evidence makes it certain that $\ulcorner\beta\urcorner$ is not true. By EVIDENCE it follows that Alpha ought to believe that $\ulcorner\beta\urcorner$ is not true. That is, we have $OB_\alpha\neg T(\ulcorner\beta\urcorner)$. As an instance of CONSISTENCY we have: $O(B_\alpha\neg T(\ulcorner\beta\urcorner) \rightarrow \neg B_\alpha T(\ulcorner\beta\urcorner))$. I assume that rational obligations are such that if a proposition γ is a consequence of a set of propositions Γ and the members of Γ are all rationally obligatory then so is γ .⁸ Given this $OB_\alpha\neg T(\ulcorner\beta\urcorner)$ follows from the previous two claims.

Cases 1 and 2 show that CONSISTENCY and EVIDENCE are *classically* inconsistent with POSSIBILITY. If Alpha does not believe that $\ulcorner\beta\urcorner$ is true, then, according to Case 1, it ought to believe it. While if Alpha does believe that $\ulcorner\beta\urcorner$ is true, then, according to Case 2, it ought to not believe it. As a theorem of classical logic we have $B_\alpha T(\ulcorner\beta\urcorner) \vee \neg B_\alpha T(\ulcorner\beta\urcorner)$. What Cases 1 and 2 show is that if CONSISTENCY and EVIDENCE hold then whichever disjunct is realized Alpha will be guilty of a rational failure. In neither case, however, will Alpha be guilty of any initial rational failing. All that we require is that Alpha have knowledge of a theorem and that it be sensitive to its own doxastic states. By proof-by-cases reasoning then we can establish that if CONSISTENCY and EVIDENCE hold it isn't possible for Alpha, an antecedently rational agent, to meet all of the requirements imposed by rationality.⁹ But this is exactly what POSSIBILITY denies.

Rejecting CONSISTENCY, EVIDENCE or POSSIBILITY, then, each brings with it significant intuitive costs. And yet the case of Agent Alpha would seem to show that we cannot accept each of these plausible normative principles. We are faced with a normative paradox.

⁸This sort of multi-premise closure principle is not completely uncontroversial. In particular, those who think that there are rational dilemmas, that is cases in which $O\phi$ and $O\neg\phi$, will want to reject such a closure principle since rational dilemmas together with this closure principle lead to deontic trivialization. For this type of worry see, for example, Van Fraassen (1973). Let me note, then, that the cases in which I will be appealing to multi-premiss closure for rational obligations are all cases in which such trivialization is avoided. So if one is inclined to be suspect of such a closure principle due to rational dilemmas there are restricted closure principles which would avoid such worries and suffice for my purposes.

⁹Note that I am assuming that on the intended reading of POSSIBILITY the modality is restricted to situations in which we hold fixed the facts about the agent's actual situation that are relevant to the rationality of particular options were they to be realized by the agent. In the case we are concerned with, then, in applying POSSIBILITY we must hold fixed the facts about Alpha that are relevant to its evidential situation. But these include all the facts that were appealed to in establishing Cases 1 and 2, namely that (3), (4), (5) and (6) all hold and that the relevant beliefs were arrived at in a particular manner. We can thus take Cases 1 and 2 for granted in applying POSSIBILITY.

3.2 Credence

The above normative paradox worked with a qualitative notion of belief. But it is quite plausible that underlying such qualitative states is a much more fine-grained quantitative doxastic structure. In addition to the question of whether or not an agent believes a certain proposition, we can ask: *how strongly* does the agent believe the proposition? An agent may believe two propositions and yet believe one more strongly than the other. Call this more fine-grained doxastic state a *credal state*. A paradox similar to that just developed can be formulated by appealing to quantitative beliefs, i.e., credences, instead of qualitative beliefs.

We'll assume that Alpha's credal state can be represented by a function mapping propositions to real numbers between 0 and 1.¹⁰ We write: $Cr(\phi) = r$, to mean that Alpha's credence in ϕ is r .

What constraints does rationality impose on Cr ? That is, what constraints, if any, does rationality impose on the values that Alpha's credences can take? Two very plausible principles are the following:

CREDENCE₁ $\models \phi \rightarrow \psi \Rightarrow O(Cr(\phi) \leq Cr(\psi))$

CREDENCE₂ $O(Cr(\neg\phi) = 1 - Cr(\phi))$

CREDENCE₁ tells us that if $\phi \rightarrow \psi$ is valid then an agent is rationally required to give at least as much credence to ψ as to ϕ . CREDENCE₂ tells us that an agent ought to be such that its credence in the negation of a proposition is always 1 - its credence in that proposition. I note that CREDENCE₁ clearly entails:

CREDENCE₃ $\models \phi \leftrightarrow \psi \Rightarrow O(Cr(\phi) = Cr(\psi))$

While there is a fair amount of controversy over the principles governing rational credences, CREDENCE₁-CREDENCE₃ are each quite plausible. We can, however, make trouble for these principles in the same way that we made trouble for CONSISTENCY and EVIDENCE. We can provide an argument that would seem to show that CREDENCE₁-CREDENCE₂ are incompatible with POSSIBILITY.

Let $\ulcorner\beta'\urcorner$ name the following sentence: $\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9$. Then as an instance of the T-schema we have:

$$(7) T(\ulcorner\beta'\urcorner) \leftrightarrow \neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9$$

Assuming that we take this instance of the T-schema to be valid then we have:

¹⁰I'm not assuming here that *any* rational agent's mental state must be able to be so represented, but just that *this particular* rational agent's mental state can be so represented. Taking credences to be represented by real numbers means that for any propositions ϕ and ψ either the agent's belief in ϕ is at least as strong as its belief in ψ , or the agent's belief in ψ at least as strong as its belief in ϕ . But one may question whether, as a normative matter, an agent's strength of belief should be required to induces such a total ordering. This is, I think, a difficult question, but it isn't one that I need to address to get the paradox up and running.

$$(8) \models T(\ulcorner \beta' \urcorner) \leftrightarrow \neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9$$

We will assume that Alpha is perfectly reliable about the credence that it has concerning the proposition that $\ulcorner \beta' \urcorner$ is true. More precisely we assume:

$$(9) \forall x \in \mathbb{R} \ Cr(Cr(T(\ulcorner \beta' \urcorner)) = x) = 1 \leftrightarrow Cr(T(\ulcorner \beta' \urcorner)) = x$$

Given that Alpha satisfies (9), we can now argue that Alpha is unable to meet the requirements imposed by CREDENCE₁-CREDENCE₂. We'll break the argument into two cases.

Case 1: On the assumption that $Cr(T(\ulcorner \beta' \urcorner)) > 0.9$, it follows that Alpha will be in violation of one of CREDENCE₁-CREDENCE₂.

To see this we argue as follows. Given (8), it follows that if Alpha meets the requirement imposed by CREDENCE₁ then: $Cr(T(\ulcorner \beta' \urcorner)) = Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. However, we can show that if Alpha meets certain other demands imposed by CREDENCE₁-CREDENCE₂ then this equality must fail. Let $x > 0.9$ and $Cr(T(\ulcorner \beta' \urcorner)) = x$. We have $\models Cr(T(\ulcorner \beta' \urcorner)) = x \rightarrow Cr(T(\ulcorner \beta' \urcorner)) > 0.9$. Assuming that Alpha meets the demands imposed by CREDENCE₁, then we have: $Cr(Cr(T(\ulcorner \beta' \urcorner)) = x) \leq Cr(Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. By (9) we have $Cr(Cr(T(\ulcorner \beta' \urcorner)) = x) = 1$. It follows that CREDENCE₁ demands that $Cr(Cr(T(\ulcorner \beta' \urcorner)) > 0.9) = 1$. Given this, CREDENCE₂ demands that $Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9) = 0$. Since $x \neq 0$, it follows that if Alpha meets certain demands imposed by CREDENCE₁-CREDENCE₂, then Alpha will fail to meet the requirement imposed by CREDENCE₁ that $Cr(T(\ulcorner \beta' \urcorner)) = Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. This shows that if $Cr(T(\ulcorner \beta' \urcorner)) > 0.9$, then Alpha will be in violation of one of CREDENCE₁-CREDENCE₂.

Case 2: On the assumption that $Cr(T(\ulcorner \beta' \urcorner)) \leq 0.9$, it follows that Alpha will be in violation of one of CREDENCE₁-CREDENCE₂.

Again, given (8), it follows that if Alpha meets the requirement imposed by CREDENCE₁ then: $Cr(T(\ulcorner \beta' \urcorner)) = Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. We can argue that if Alpha meets certain other demands imposed by CREDENCE₁ then this equality must fail. Let $y \leq 0.9$ and $Cr(T(\ulcorner \beta' \urcorner)) = y$. We have $\models Cr(T(\ulcorner \beta' \urcorner)) = y \rightarrow \neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9$. Assuming that Alpha meets the demands imposed by CREDENCE₁, then we have: $Cr(Cr(T(\ulcorner \beta' \urcorner)) = y) \leq Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. By (8) we have $Cr(Cr(T(\ulcorner \beta' \urcorner)) = y) = 1$. It follows that CREDENCE₁ demands that $Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9) = 1$. Since $y \neq 1$, it follows that if Alpha meets certain demands imposed by CREDENCE₁, then Alpha will fail to meet the requirement imposed by CREDENCE₁ that $Cr(T(\ulcorner \beta' \urcorner)) = Cr(\neg Cr(T(\ulcorner \beta' \urcorner)) > 0.9)$. This shows that if $Cr(T(\ulcorner \beta' \urcorner)) \leq 0.9$, then Alpha will be in violation of one of CREDENCE₁, and so *a fortiori* one of CREDENCE₁ - CREDENCE₂.

What Cases 1 and 2 would seem to tell us is that if Alpha has perfect access to its credence in the proposition that $\ulcorner\beta\urcorner$ is true, it follows that Alpha is unable to meet the requirements imposed by CREDENCE_1 - CREDENCE_2 .¹¹ Since this type of access to one's credal state need not in itself involve any failure of rationality, what Cases 1 and 2 would seem to tell us is that CREDENCE_1 - CREDENCE_2 are incompatible with POSSIBILITY . For if CREDENCE_1 - CREDENCE_2 are both true, then, it would seem to be possible for an antecedently rational agent to be doomed to irrationality. But this is exactly what POSSIBILITY denies. We once again are faced with a normative paradox.

3.3 Knowledge

Here is a final paradox.¹² This case concerns states of knowledge.

Here are two very plausible principles concerning knowledge:

FACTIVITY ϕ is a consequence of the claim that one knows ϕ .¹³

DEDUCTION If one knowingly deduces ϕ by a known valid method, then one knows ϕ .

Despite the plausibility of these two principles we can make trouble for them as follows.

Let K stand for *Alpha knows that*. Let $\ulcorner\kappa\urcorner$ name the following sentence: $\neg KT(\ulcorner\kappa\urcorner)$. So as an instance of the T-schema we have:

$$(10) T(\ulcorner\kappa\urcorner) \leftrightarrow \neg KT(\ulcorner\kappa\urcorner)$$

Now consider the following simple derivation:

1	$KT(\ulcorner\kappa\urcorner)$	
2	$T(\ulcorner\kappa\urcorner)$	1, FACTIVITY
3	$\neg KT(\ulcorner\kappa\urcorner)$	2, (10), \leftrightarrow -Elim
4	$\neg\neg KT(\ulcorner\kappa\urcorner)$	1-3 \neg -Intro
5	$T(\ulcorner\kappa\urcorner)$	4, (10) \leftrightarrow -Elim

¹¹It is perhaps worth noting one way in which we could weaken our principles and preserve the above paradox. The only place in which appeal is made to CREDENCE_2 is in Case 1. And there we could get the same result with the weaker assumption: $O(Cr(\neg\phi) \leq 1 - Cr(\phi))$. One reason that this is worth noting is that while proponents of standard Bayesian theories will accept CREDENCE_2 , there are alternative accounts of the norms governing credence, notably those that hold that credences should be representable by Dempster-Shafer functions, that give up CREDENCE_2 in favor of this weaker constraint. Our paradox, then, has force even for these non-Bayesians.

I note also that the fact that our paradox can be generated with the weaker principle will prove relevant later when we turn to the question of the cognitive significance of non-classical semantic values.

¹²This paradox was discussed in Kaplan and Montague (1960).

¹³Read: $K\phi \models \phi$.

It would seem that there is a simple derivation of the claim that $\ulcorner\kappa\urcorner$ is true. But then this seems like a fact that Alpha could know. And so Alpha should be able to knowingly deduce the claim that $\ulcorner\kappa\urcorner$ is true. But then it would follow by DEDUCTION that Alpha knows that $\ulcorner\kappa\urcorner$ is true, i.e., $KT(\ulcorner\kappa\urcorner)$. But we've already proven that Alpha doesn't know that $\ulcorner\kappa\urcorner$ is true, i.e., $\neg KT(\ulcorner\kappa\urcorner)$. FACTIVITY and DEDUCTION, though each individually quite plausible would seem to lead to a contradiction. Given classical logic, it would seem that we can't accept both.

3.4 Some Responses

The paradoxes just considered have been familiar to philosophers for some time now, and various responses have been suggested. In the next chapter, we'll see how we can resolve these paradoxes by appealing to the non-classical logics surveyed in chapter 2. What I'll claim is that *if* one accepts such a non-classical logic, then one should resolve the paradoxes by appealing to the appropriate non-classical resources. A large part of the motivation for this response comes from seeing how difficult it is to provide a well-motivated response to these paradoxes, given the acceptance of classical logic. Let us turn, then, to consider the responses available given this constraint.

3.4.1 Responses to the Paradox of Belief

In §3.1, we saw that EVIDENCE, CONSISTENCY and POSSIBILITY led to a contradiction, on the assumption that it was possible for there to be an antecedently rational agent, Alpha, who believed the theorem $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$ and who had perfect access to whether or not it believed that $\ulcorner\beta\urcorner$ is true. The obvious strategies for resolving the paradox are:

- Reject EVIDENCE
- Reject CONSISTENCY
- Reject POSSIBILITY
- Claim that there is something *impossible* or *antecedently irrational* about a putative agent such as Alpha.

Let us take these in turn.

Rejecting EVIDENCE

It is certainly *prima facie* quite plausible that there is a level of evidence that rationally mandates belief. Nonetheless, it is tempting to think that what the paradox of belief shows us is that this thought is ultimately incorrect.

The paradox of belief crucially exploits the fact that there can be situations in which an agent has evidence that makes a proposition certain, but the agent's having this evidence depends on the

agent *not* responding by believing the proposition in question. A natural thought is that at least in those cases in which one's having a certain body of evidence depends on one's not responding in the typically appropriate way, such evidence won't have its normal normative force. This type of response to the paradox of belief is advocated, for example, in Conee (1987).

One way of developing this idea would be to maintain that it isn't evidence that has a rational bearing on whether one should believe ϕ , but instead it is one's evidence *conditional on the supposition that one believes ϕ* .¹⁴

Instead of EVIDENCE, the thought would be that we should accept:

CONDITIONAL EVIDENCE For any proposition ϕ , if, conditional on the supposition that the agent believes ϕ , an agent's evidence makes ϕ certain, then the agent is rationally required to believe ϕ .

It is not hard to see how trading in CONDITIONAL EVIDENCE for EVIDENCE allows us to block the paradox presented in §3.1. Case 1 relied on the fact that when Alpha does not believe that $\ulcorner\beta\urcorner$ is true, it will be in a position in which its evidence makes it certain that $\ulcorner\beta\urcorner$ is not true, and so, given EVIDENCE, it will be in violation of a normative requirement. The key point to note here is that although in this situation Alpha's evidence makes it certain that $\ulcorner\beta\urcorner$ is true, *conditional* on the assumption that Alpha believes that $\ulcorner\beta\urcorner$ is true, Alpha's evidence will make it certain that $\ulcorner\beta\urcorner$ is *not* true. And so if we trade in EVIDENCE for CONDITIONAL EVIDENCE, we can block the argument that Alpha will be in violation of a rational requirement given that it does not believe that $\ulcorner\beta\urcorner$ is true.

Despite the elegance of this response, I don't think that it ultimately gets to the bottom of this paradox. And the reason for this is that we can develop an analogous paradox that does not rely at all on the principle EVIDENCE; i.e., we can whittle the apparent inconsistency down to the pair CONSISTENCY and POSSIBILITY. Here's how that works.

Belief, it is common to assume, is a relation that holds between an agent and an abstract object, a proposition. Assuming this picture of belief we can show that CONSISTENCY and POSSIBILITY are classically inconsistent. Consider the following propositional analogue of $\ulcorner\beta\urcorner$. Let (*) name the following sentence:

Alpha doesn't believe the proposition expressed by (*)¹⁵

Let's abbreviate *the proposition expressed by* as ρ . The above can, then, be represented as:

(*) $\neg B_\alpha \rho(*)$

Note that since both (*) and ' $\neg B_\alpha \rho(*)$ ' name the same sentence the following holds:

(r1) $\rho(*) = \rho'\neg B_\alpha \rho(*)'$

¹⁴This idea is argued for on independent grounds in Wedgwood (2008).

¹⁵Again we should think of *Alpha* as having a first-personal mode of presentation for Alpha.

Our transparency assumptions can be captured by the following analogues of (3) and (4):

$$(11) B_{\alpha}\rho(*) \leftrightarrow B_{\alpha}\rho' B_{\alpha}\rho(*)'$$

$$(12) \neg B_{\alpha}\rho(*) \leftrightarrow B_{\alpha}\rho' \neg B_{\alpha}\rho(*)'$$

It can now easily be shown that the assumption that Alpha does not believe the proposition expressed by (*) leads to a contradiction.

1	$\neg B_{\alpha}\rho(*)$	
2	$B_{\alpha}\rho' \neg B_{\alpha}\rho(*)'$	1, (12)
3	$B_{\alpha}\rho(*)$	2, (r1)
4	$B_{\alpha}\rho(*)$	1-3

It follows classically that Alpha cannot fail to believe the proposition expressed by (*). However, when Alpha believes this proposition, given (11), it is doomed to inconsistency. Thus:

1	$B_{\alpha}\rho(*)$	
2	$B_{\alpha}\rho' B_{\alpha}\rho(*)'$	1, (11)
3	$B_{\alpha}\rho' \neg B_{\alpha}\rho(*)'$	1, (r1)

Given CONSISTENCY, we will once again have a violation of POSSIBILITY. Holding fixed (11) and (12), it follows that Alpha's only option is to believe the proposition expressed by (*). But in doing so Alpha will be in violation of CONSISTENCY. It follows that it is not possible for Alpha to meet the rational requirements imposed by CONSISTENCY. Since Alpha need not be guilty of any antecedent rational failing, this is a violation of POSSIBILITY.

Let me say a little about how this case is related to our earlier case. The key difference is the replacement of $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_{\alpha}T(\ulcorner\beta\urcorner)$ by $\rho(*) = \rho' \neg B_{\alpha}\rho(*)'$. Changing a conditional linking the truth-values of propositions to an identity between propositions has the same effect as assuming conformity to EVIDENCE. If we assume that Alpha meets EVIDENCE we can provide parallel derivations to those involving the sentence $\ulcorner\beta\urcorner$.

Corresponding to our first derivation we have:

1	$\neg B_\alpha T(\ulcorner\beta\urcorner)$	
2	$B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)$	1, (4)
3	$B_\alpha T(\ulcorner\beta\urcorner)$	2, (6), EVIDENCE
4	$B_\alpha T(\ulcorner\beta\urcorner)$	1-3

Corresponding to our second derivation we have:

1	$B_\alpha T(\ulcorner\beta\urcorner)$	
2	$B_\alpha B_\alpha T(\ulcorner\beta\urcorner)$	1, (3)
3	$B_\alpha \neg T(\ulcorner\beta\urcorner)$	2, (5), EVIDENCE

Where in the former derivations appeal is made to the propositional identity $\rho(*) = \rho'\neg B_\alpha\rho(*)$, in the latter we must appeal to Alpha's justified belief in $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$, together with the assumption that Alpha meets the evidential norm EVIDENCE. Appeal to propositions such as that expressed by (*) obviates the need for an appeal to EVIDENCE. The conflict between CONSISTENCY, EVIDENCE and POSSIBILITY can thereby be reduced to a conflict between CONSISTENCY and POSSIBILITY.

One thing that may be worrisome about this paradox is the appeal to propositions. Why should we assume that (*) does in fact express a proposition which could serve as the object of Alpha's belief? I take it that the worry here is the self-referential nature of (*). In response to this worry let me make the following observations.

First we should fix on some diagnostic tests for whether a sentence ϕ expresses a proposition. I take it that a sufficient condition for ϕ to express a proposition is if ϕ can be embedded under metaphysical or doxastic operators in a way that results in a true sentence. For the resultant sentence could be true only if it expressed a proposition; and such a sentence could express a proposition only if its component sentences expressed propositions. A sentence's failure to express a proposition is something which infects any sentence of which it is a part.

Given this, we can show that a sentence is not, in general, precluded from expressing a proposition in virtue of the fact that it contains a term that purports to refer to the proposition expressed by that sentence.

One way to achieve sentential self-reference is via stipulation, as in the case of (*). Another is via a definite description that picks out the sentence in which the definite description occurs.¹⁶ Imagine, for example, that in room 301 there is a single blackboard, and on that blackboard is written the following sentence: 'The proposition expressed by the sentence on the blackboard in

¹⁶This is noted (perhaps for the first time) in Kripke (1975).

room 301 is not true.’ In this case, the definite description: ‘the sentence on the blackboard in room 301’, refers to the very sentence of which that definite description is a constituent. And so the definite description: ‘the proposition expressed by the sentence on the blackboard in room 301’ purports to refer to the proposition expressed by that sentence.

Let us ask whether this sentence does indeed express a proposition. To argue that the answer is ‘yes’ it suffices to argue that this sentence can embed under metaphysical and doxastic operators in a way that results in a true sentence.

It seems fairly obvious that this sentence can embed under doxastic operators and yield a true sentence. For example, let John be someone who believes that there is just one sentence written on the blackboard in 301 and that that sentence is: ‘ $2 + 2 = 5$ ’. Let John further believe that the proposition expressed by the sentence written on the blackboard in 301 is the proposition that $2 + 2 = 5$ and that this is not true. Given these beliefs it would seem that John believes that the proposition expressed by the sentence written on the blackboard in 301 is not true.¹⁷ It would seem, then, that we can perfectly well embed: ‘The proposition expressed by the sentence on the blackboard in room 301 is not true.’, under the operator ‘John believes that...’ and get a true sentence. But if that’s the case then it must be that ‘The proposition expressed by the sentence on the blackboard in room 301 is not true.’ expresses a proposition.

Similarly, it seems clear that this sentence can embed under metaphysical modal operators and produce a true sentence. Consider a possible world in which the sentence written on the blackboard in 301 is ‘ $2 + 2 = 5$ ’. We assume that in this world the proposition expressed by the sentence written on the blackboard in 301 is just the proposition that $2 + 2 = 5$. In this world, then, the proposition expressed by the sentence written on the blackboard in room 301 is not true. But then it follows that it is possible that the proposition expressed by the sentence written on the blackboard in room 301 is not true.¹⁸ It would seem, then, that we can embed: ‘The proposition expressed by the sentence on the blackboard in room 301 is not true.’, under the operator ‘It is possible that...’ and get a true sentence. And if that’s the case, then ‘The proposition expressed by the sentence on the blackboard in room 301 is not true’ must express a proposition.

The above reflections show that a sentence is not barred from expressing a proposition simply in virtue of containing a singular term that purports to refer to the proposition expressed by that sentence. Given that this is the case, it is hard to see what good reason there could be to deny that (*) fails to express a proposition.

What the above paradox shows is that rejecting EVIDENCE will not ultimately resolve the normative puzzle raised by the paradox in §3.1

Of course, that doesn’t mean that we should accept EVIDENCE. Indeed, one may still think that it is rather plausible that evidence can’t rationally mandate belief where that evidence is of a sort to which one cannot respond. And, given this sort of reflection, one might be inclined towards a principle like CONDITIONAL EVIDENCE. Now, ultimately I think it is a very difficult question whether one should accept EVIDENCE or CONDITIONAL EVIDENCE. It is worth, however,

¹⁷Perhaps this doesn’t follow from the above assumptions. Perhaps there is room here for John to have the former beliefs but lack the latter. Nonetheless given that John has the relevant conceptual capacities it surely perfectly reasonable to assume that John has this belief, given that he has the former.

¹⁸Of course, this is only true on the *de dicto* reading of the above sentence.

noting a way in which our trimmed down paradox would seem to make trouble for CONDITIONAL EVIDENCE.

The general thought behind CONDITIONAL EVIDENCE, I take it, is that insofar as we are concerned with having true beliefs what is of relevance is not whether or not something is true, but whether it is true *given that we believe it*. And so what should be guiding our beliefs is not our evidence (which tells us about what is true), but our evidence *conditional on our believing certain propositions* (which tells us about what is true given that we believe certain propositions). Given this motivation, the proponent of CONDITIONAL EVIDENCE should also accept the following principle:

CONDITIONAL EVIDENCE* For any proposition ϕ , if an agent's evidence makes it certain that ϕ is false, conditional on the supposition that the agent believes ϕ , then the agent is rationally required to not believe ϕ .

The problem is the following. Assuming (as we can) that Alpha knows that the proposition expressed by (*) is true just in case Alpha doesn't believe the proposition expressed by (*), it follows that conditional on Alpha's believing this proposition, Alpha will be certain that the proposition is false. Given CONDITIONAL EVIDENCE*, it follows that Alpha will be required not to believe the proposition expressed by (*). But as we have seen, given its introspective powers, Alpha is guaranteed to believe this proposition. And so Alpha is guaranteed to be in violation of a requirement imposed by CONDITIONAL EVIDENCE*. This would seem to show that CONDITIONAL EVIDENCE* is incompatible with POSSIBILITY.

Now in principle one could hold on to CONDITIONAL EVIDENCE while rejecting CONDITIONAL EVIDENCE*, but it is hard to see what principled reason there could be for such seemingly invidious treatment. The problem for CONDITIONAL EVIDENCE* would also seem to be a problem for CONDITIONAL EVIDENCE. This, of course, doesn't mean that we should reject CONDITIONAL EVIDENCE. But what it does mean, I think, is that we shouldn't prefer EVIDENCE to CONDITIONAL EVIDENCE, on the grounds that EVIDENCE leads to situations in which an agent cannot respond to the normative demands imposed by the constraint.

Rejecting CONSISTENCY

What made rejecting EVIDENCE at least *prima facie* attractive was that there seemed to be available a plausible alternative principle, viz., CONDITIONAL EVIDENCE, that could account for the former principle's intuitive appeal. The problem with rejecting EVIDENCE, however, was that although it allowed us to block the paradox developed in §3.1, it was insufficient to deal with a very closely related paradox.

The situation with the rejection of CONSISTENCY is, in a certain sense, the reverse. If we reject CONSISTENCY, then we can block both paradoxes. However, it is very hard to provide a plausible story to explain *why* it is that CONSISTENCY fails. What we would like, ideally, is a non-ad hoc alternative principle that delivers the same results as CONSISTENCY in standard cases, but allows for the relevant exceptions. Simply stating that CONSISTENCY is correct insofar as it

doesn't lead to normative paradox is not, I think, sufficient. But what more detailed and principled story one should tell is far from obvious.

Here's it's worth noting that although one may want to allow that a rational agent's set of beliefs need not be consistent, the standard motivations for this do nothing to motivate the idea that it could be rational to believe both a proposition and its negation.

For example, the preface paradox provides a pretty convincing case in which it would seem to be rational for one to believe some series of claims C_1, C_2, \dots, C_n while believing the negation of their conjunction $\neg(C_1 \wedge C_2 \wedge \dots \wedge C_n)$.¹⁹ This sort of case, however, provides no reason to think that there are contradictory pairs of propositions which it might be rational to believe. The reason for this is that the best explanation for why it could be rational to believe all of C_1, C_2, \dots, C_n and $\neg(C_1 \wedge C_2 \wedge \dots \wedge C_n)$, is that there is some high level of evidential support which suffices for a belief to be rational. Given a large enough number n , C_1, C_2, \dots, C_n could all have extremely high evidential support, while their conjunction would have a very low evidential support. The levels of evidential support for a claim and its negation are inversely correlated; any increase in support for one will be a decrease in support for the other (at least taking classical logic for granted). Given, then, that $C_1 \wedge C_2 \wedge \dots \wedge C_n$ has very low evidential support, $\neg(C_1 \wedge C_2 \wedge \dots \wedge C_n)$ will have very high evidential support, and so will be rational to believe.

But clearly if this is the reason that we think that it may in certain cases be rational to have inconsistent beliefs, we have no reason to think that it may be rational to have *pairwise* inconsistent beliefs. As long as we take the level of evidential support which rationalizes a belief to be greater than that which is evenly balance between supporting the truth and falsity of a claim, then, given the inverse correlation between the support a claim receives and that which its negation receives, we need never say that it is rational to believe a claim ϕ and its negation $\neg\phi$. And it seems absurd to think that the evidence for a claim being evenly balanced between supporting the truth and the falsity of the claim is sufficient for the belief in that claim to be rational. The correct attitude in such a situation is agnosticism. So, while we may have reason to give up consistency of belief as a general requirement, the cases which motivate our so doing provide no reason to think that we should also give up the constraint of pairwise consistency.

Of course, all of this simply amounts to a challenge to one who wants to defend the rejection of CONSISTENCY. The challenge is to provide some plausible story that will explain why it can be rational to have pairwise inconsistent beliefs. Perhaps there is such a story to be offered, but I'm skeptical.

Rejecting POSSIBILITY

Given the problems with rejecting EVIDENCE and CONSISTENCY, one might be tempted to think that the problem ultimately lies with POSSIBILITY. Indeed the idea that normative demands must always in principle be capable of being met is one which has been thought to be problematic in other normative domains.

For example in the domain of practical rationality there would appear to be certain cases in

¹⁹See Makinson (1965).

which whatever choice an agent makes the agent ought to not have chosen that option. As an example, consider the case of Death in Damascus in Gibbard and Harper (1978). In this situation an agent is choosing between staying in Damascus or going to Aleppo. The agent believes that, conditional on his being in Damascus, Death will certainly visit him there, while, conditional on his being in Aleppo, Death will certainly visit him there. In this case, it seems that whichever decision the agent makes—whether the agent decides to stay in Damascus or decides to go to Aleppo—the agent will be certain that things would have been better had he made the other choice. It's important here to note that although the agent is certain that Death will meet him in Damascus (Aleppo) conditional on the assumption that the agent is in Damascus (Aleppo) tomorrow, the agent doesn't think that there is any causal connection between what the agent decides to do. So once the agent decides, say, to stay in Damascus, and so becomes certain that Death will meet him in Damascus, the agent will become certain that had he instead decided to go to Aleppo he would have avoided death. In this type of situation standard causal decision theory will hold that if the agent decides to stay in Damascus then the agent ought to have decided to go to Aleppo, while if the agent decides to go to Aleppo then the agent ought to have decided to stay in Damascus. It would seem, then, that if one endorses causal decision theory, there is pressure to give up POSSIBILITY as a constraint on principles of rationality.²⁰

We should not, however, underestimate the intuitive costs of this response. It is, I think, initially quite implausible that an agent could do everything that rationality requires and yet nonetheless wind up in a situation in which it cannot continue to meet the requirements of rationality.

This intuition can be bolstered by considering the sorts of conditions under which rational criticism seems to be appropriate. Let's focus on the case of doxastic rationality. An agent may be subject to rational criticism given the set of doxastic options that it has realized. Let Γ be this set. The following seems to me to be a plausible constraint on the conditions under which such criticism is appropriate:

APPROPRIATENESS If an agent is to be subject to rational criticism for realizing Γ , then there is some set of sets of options Δ meeting the following conditions:

- (i) Each member of Δ is incompatible with Γ .
- (ii) The agent should have realized some member of Δ (although there need not be any particular member that it should have realized).
- (iii) Each member of Δ is such that had the agent realized this set of doxastic options rational criticism would have been inappropriate.

Justification: If an agent is subject to rational criticism given the total set of doxastic options, Γ , that it has realized, then it should not have realized Γ . In such a case the agent ought to have realized some other set of doxastic options (although there need not be a specific set of options

²⁰Pressure also comes from the moral domain. See, for example, Lemmon (1962) and Marcus (1980) for arguments that moral dilemmas are simply a fact of life. See also Priest (2002) for a general argument in favor of rational dilemmas.

that the agent ought to have realized). That is, there will be a set Δ of sets of options incompatible with Γ such that the agent ought to have realized one of the members of Δ . (Indeed, there will typically be many such sets.) That the agent ought to have realized one of the members of Δ can serve as the grounds for rational criticism for the agent's having instead realized Γ . If, however, failure to realize some member of Δ is to serve as an adequate ground for rational criticism, it must, I think, be the case that the agent's realizing some member of Δ would have made rational criticism inappropriate. Rational criticism for an agent's doxastic situation is grounded in the idea that the agent should be some other way that would have made such criticism inappropriate. It is this plausible intuition that APPROPRIATENESS captures.

It can be shown that if POSSIBILITY fails then so must APPROPRIATENESS.

Here's the argument for this claim: Let Γ pick out the set of doxastic options that a rationally blameless agent has realized. Assume that POSSIBILITY fails. This means that there must be some jointly exhaustive set of options Σ such that for every $\sigma' \in \Sigma$ the agent is rationally culpable if it realizes $\Gamma \cup \sigma'$. We pick some arbitrary member σ of Σ . Let Δ be an arbitrary set of sets of options incompatible with $\Gamma \cup \sigma$, such that the agent should realize one of these sets. I'll argue that there are members of Δ such that were an agent to realize that option then it would be rationally culpable. This shows that a violation of POSSIBILITY leads to a violation of APPROPRIATENESS.

Amongst the members of Δ must be some set containing Γ , since if one ought to realize some set amongst a collection of sets all of which are incompatible with Γ , then one would, contrary to hypothesis, be rationally blameworthy in realizing Γ . However, since Γ is not itself incompatible with $\Gamma \cup \sigma$, then any set in Δ containing Γ must also contain some other doxastic option(s) Γ' , in addition to those options in Γ . By hypothesis, $\Gamma \cup \Gamma'$ is incompatible with $\Gamma \cup \sigma$, i.e., $\Gamma \cup \Gamma' \cup \sigma \models \perp$. It follows that $\Gamma \cup \Gamma' \models \neg\sigma$. But given this, our agent will be rationally culpable in realizing $\Gamma \cup \Gamma'$, since this will involve realizing $\Gamma \cup \neg\sigma$, and so realizing Γ together with some member of $\Sigma - \sigma$.

If, then, one wants to maintain what is, I think, a natural principle about the conditions for appropriate rational criticism then one should endorse POSSIBILITY.²¹

Given the plausibility of POSSIBILITY, there have been attempts to try to salvage this principle and make it compatible with a plausible decision theory. According to this line of response what we need to do in order to satisfy the demands imposed by POSSIBILITY is to expand the space of possible options available to the agent. In addition to the options of staying in Damascus and going to Aleppo we need to allow for so-called "mixed-decisions", which are formed by probabilistic weightings between these two options.²² Now, this would not be the place to consider such responses in any detail. But it is worth mentioning them, for the responses that I'll be advocating on behalf of various proponents of non-classical logic have something in common with this type of response. The idea will be to allow for mental states of a kind not considered in the framing of the putative rational dilemma. I take it that this type of response is well motivated given the plausibility of POSSIBILITY.

²¹See Conee (1982) for further grounds for rejecting the possibility of rational dilemmas.

²²See e.g., Osborne and Rubenstein (1994) for an overview of standard game-theoretic treatment of mixed-decisions. See, also, Arntzenius (2008) for a very good discussion of how we can make sense of these options and the role that they play in rational decision making.

Rejecting The Set-Up

Another option would be to reject one of the following elements of the set-up of the case:

- (i) Reject the claim that an agent such as Alpha may rationally believe: $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$.
- (ii) Reject the claim that there is some possible antecedently rational agent with the introspective powers assigned to Alpha.

I'll put off discussion of option (ii) until we discuss possible responses to the credential paradox.

Option (i) has been surprisingly popular, and though I think that it ultimately has little to recommend for itself, given that it has been endorsed by other authors in response to similar paradoxes, it is worth discussing. Call an *anti-expert* about some proposition ϕ someone who is such that ϕ is true just in case the agent does not believe ϕ .²³ Various authors (e.g., Sorensen (1988), Elga and Egan (2005)) have argued that no agent can rationally take themselves to be an anti-expert about a proposition ϕ . The argument offered for this claim is essentially that if such an agent does take itself to be an anti-expert and is aware of its own mental states then the agent will be in violation of various other rational requirements.

One problem with this response is that if an agent is an anti-expert about some proposition then it would seem to be at least possible for the agent to have evidence that makes this fact certain. For example, the agent could be told by various trusted informants, or in the case we considered, since $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$ is valid, the agent could go through the relevant reasoning that establishes the claim of anti-expertise. Claiming that rationality requires the agent to not self-ascribe anti-expertise amounts, in this situation, to the claim that rationality requires an agent to ignore its evidence. Note that this is stronger than simply claiming that we should reject the principle EVIDENCE. Call a view of rationality that holds that rationality can demand of an agent that it ignore evidence *ostrich rationality*. A problem with option (i) is that *ostrich rationality* just doesn't seem to be a form of rationality at all.

Of course, this is simply to note the *prima facie* implausibility of this response. If, in the end, option (i) offered a solution to our paradox then perhaps we should revise our initial impressions. We can, however, show that option (i) is insufficient to adequately deal with our paradox. The problem is the same as that facing the rejection of EVIDENCE. Although option (i) will block the paradox developed in §3.1, it does nothing to solve the alternative propositional version of the paradox. The reason for this is that at no point in the development of the latter paradox did we assume that Alpha believed that it is an anti-expert. Option (i), then, can't ultimately provide a fully adequate resolution of our normative paradox. And so, given its initial implausibility, we should reject the claim that an agent is always rationally precluded from self-ascribing anti-expertise.

²³The terminology here derives from Sorensen (1988). Actually what I'm calling anti-expertise is what Sorensen calls *omissive anti-expertise*.

3.4.2 Responses to the Paradox of Credence

In §3.2 we saw that CREDENCE_1 - CREDENCE_2 were jointly incompatible with POSSIBILITY , on the assumption that there is an antecedently rational agent Alpha who has perfect access to its credence in the proposition that $\lceil \beta \rceil$ is true. In response to this paradox the natural options are:

- Reject POSSIBILITY
- Reject one of CREDENCE_1 - CREDENCE_2
- Reject the set-up, i.e., the claim that it is possible for there to be an antecedently rational agent such as Alpha.

We've already considered the option of rejecting POSSIBILITY . Let's consider the other two options.

Rejecting one of CREDENCE_1 - CREDENCE_2

Recall that we are, at this point, considering the types of response that are available if one thinks that propositions about an agent's doxastic state obey classical logic. Now, if one takes classical logic for granted, there are a number of powerful arguments for CREDENCE_1 and CREDENCE_2 . Since the instances of CREDENCE_1 and CREDENCE_2 in §3.2 only involved propositions about an agent's doxastic state, these arguments, then, create serious problems for one who wants to (a) reject the appeal to CREDENCE_1 or CREDENCE_2 in the paradox developed in §3.2, and (b) hold that propositions about an agent's doxastic state obey classical logic.

The best developed arguments for CREDENCE_1 and CREDENCE_2 are all instances of more general argumentative strategies designed to show that rational agent's credences must be *probabilistically coherent*. If we assume that classical logic holds, then CREDENCE_1 and CREDENCE_2 provide requirements that any probabilistically coherent belief state will meet.²⁴ The arguments that have been offered for probabilism, can, therefore, be used to argue that if one accepts classical logic for propositions about an agent's doxastic state, then one should accept that CREDENCE_1 and CREDENCE_2 provide rational constraints at least for beliefs involving such propositions. While in principle the paradox developed in §3.2 can be blocked by rejecting the appropriate instances of CREDENCE_1 or CREDENCE_2 , for this option to be at all plausible we would need a story about where the arguments in favor of probabilism go wrong.

It would take us too far afield to consider in any serious detail the various arguments that have been taken to show that rational credences must be probabilistically coherent. Let me, however, briefly sketch two of the most interesting such arguments.

One way of arguing for probabilism is by appeal to so-called Dutch Book arguments. Such an argument works as follows.

²⁴The key point to note here is that if we assume that classical logic holds, then \rightarrow is equivalent to the material conditional. CREDENCE_1 then becomes the claim that if $\models \phi \supset \psi$ then a rational agent will be such that $Cr(\phi) \leq Cr(\psi)$. And this condition is satisfied if one is probabilistically coherent.

First, it is assumed that, at least for an ideal agent such as Alpha, the following connection holds between the agent's credences and its betting behavior: If $Cr(\phi) = x$, then the agent will deem as fair a bet offering $\$z$, if ϕ , for the price $\$x \times z$. To say that the agent deems such a bet as fair is to say that the agent would be willing to either buy the bet for $\$x \times z$ or sell the bet for this same amount.

Second, given this assumed connection between credence and betting behavior, it is shown that if an agent's credences are not probabilistically coherent, then the agent will deem as fair a set of bets that guarantee a loss. See Ramsey (1931) and de Finetti (1974) for the original versions of this argument.

Given that an agent is probabilistically coherent only if the agent satisfies the requirements imposed by $CREDENCE_1$ and $CREDENCE_2$, it follows that if an agent violates $CREDENCE_1$ or $CREDENCE_2$ then the agent will be subject to a Dutch Book.

Dutch Book Argument for $CREDENCE_1$: Assume that $\models \phi \rightarrow \psi$ and that $Cr(\phi) > Cr(\psi)$. We'll show that Alpha will deem fair a set of bets that guarantee it a sure loss.

Bet 1: Costs $\$Cr(\phi)$, pays $\$1$ if ϕ

Bet 2: Costs $\$Cr(\psi)$, pays $\$1$ if ψ

To guarantee Alpha a sure loss, we sell Alpha bet 1 and buy from Alpha bet 2. Having agreed to these bets, we will have $\$Cr(\phi) - Cr(\psi)$, which is guaranteed to be positive since we're assuming that $Cr(\phi) > Cr(\psi)$. At the beginning of the bet, then, we will have gained money from Alpha. However, since $\models \phi \rightarrow \psi$, we can be assured that there will be no corresponding gain for Alpha to make at the conclusion of the bets. Obviously, if we have $\neg\phi$, then Alpha will receive no extra money. And if we have ϕ , then, while Alpha will receive $\$1$ from bet 1, Alpha will also lose $\$1$ from bet 2, leaving Alpha with a net loss.

Dutch Book Argument for $CREDENCE_2$: Assume that $Cr(\neg\phi) \neq 1 - Cr(\phi)$. We'll show that Alpha will deem fair a set of bets that guarantee it a sure loss.

Bet 1: Costs $\$Cr(\neg\phi)$, pays $\$1$ if $\neg\phi$

Bet 2: Costs $\$Cr(\phi)$, pays $\$1$ if ϕ

We consider two cases.

First, assume that $Cr(\neg\phi) < 1 - Cr(\phi)$. In this case, in order to guarantee Alpha a sure loss we simply buy both bets from Alpha. In this case we will have paid $\$Cr(\phi) + Cr(\neg\phi)$. Since we're assuming $Cr(\neg\phi) < 1 - Cr(\phi)$, we know that $Cr(\phi) + Cr(\neg\phi) < 1$. What this tells us is that whether or not ϕ holds, we will be guaranteed to make money, since in either case Alpha will need to pay us more than we paid for the combined package of bets.

Next, assume that $Cr(\neg\phi) > 1 - Cr(\phi)$. In this case instead of buying bets 1 and 2 from Alpha, we sell Alpha these bets. Since we're assuming $Cr(\neg\phi) > 1 - Cr(\phi)$, we know that $Cr(\phi) + Cr(\neg\phi) > 1$. This guarantees that whether or not ϕ occurs we will make money, since in either case we will have to pay less than Alpha paid us for the combined package of bets.

What the above two arguments show is that unless Alpha conforms to CREDENCE₁ and CREDENCE₂, Alpha will deem fair a set of bets that will guarantee it a loss. Does this show that CREDENCE₁ and CREDENCE₂ provide genuine constraints on epistemic rationality?

This is a tricky question. There certainly would seem to be *something* wrong with being in a situation in which one was subject to a sure loss book of bets. But, one might worry that insofar as the agent is guilty of some failure of rationality in virtue of being subject to such a set of bets, this is a matter, at least in the first instance, of the agent being *practically* irrational. And, it doesn't, in general, follow from the fact that an agent would be practically irrational to have a certain doxastic state, that the agent would also be epistemically irrational. Practical and epistemic rationality can come apart. So, for example, you may have evidence that makes it certain that your best friend has been defrauding you. As a matter of epistemic rationality, then, you would seem to be in a situation where you ought to believe that your best friend has been defrauding you. It may be, however, that believing this would bring horrible practical consequences. It would ruin a relationship that you value for myriad reasons, while, let's say, remaining in ignorance would simply cost you some relatively small sum of money. In this case, it would seem that, as a matter of practical rationality, you ought not believe that your friend has been defrauding you. So what is mandated as a matter of practical rationality may be different than what is mandated as a matter of epistemic rationality. Given the potential gap between the verdicts of practical and epistemic rationality, we can't, then, simply infer from the fact that Alpha ought to, as a matter of practical rationality, conform to the strictures imposed by CREDENCE₁ and CREDENCE₂, that Alpha ought to conform to these strictures as a matter of epistemic rationality.

The question, then, is whether our Dutch Book arguments provide evidence for an agent's epistemic irrationality that goes beyond the practical consequences of violating CREDENCE₁ or CREDENCE₂? Here I'm tempted to say 'yes'. Plausibly what is happening in the case of Dutch Book arguments is that the agent is failing to appreciate connections between the outcomes of the bets that are *a priori* accessible. The susceptibility in such case, then, to sure-loss bets would seem to be the result of an underlying cognitive failure. Given this, it seems reasonable to take such cases to provide at least *prima facie* evidence that an agent is epistemically irrational. The above Dutch Book arguments, then, provide *prima facie* reason to take CREDENCE₁ and CREDENCE₂ seriously as putative principles governing epistemic rationality.

It would be nice, however, if we had an argument for CREDENCE₁ and CREDENCE₂ that was able to screen off the noise of practical irrationality. In response to such worries, attempts have been made more recently to argue that probabilism is forced on us just by considering purely epistemic goods.

Truth, it is common to think, is the goal of belief. Now exactly what this amounts to is a difficult question, but one attractive thought is that truth is a primary epistemic good. Whatever else we may want to say about a belief, if it is true, then from an epistemic perspective something

has gone right. Now, when we turn to credences, we can't assess an agent's credence in a proposition as being true or false simpliciter. We can, however, assess whether such a credence is *close* to the truth or not. If ϕ is true, then the higher the agent's credence in ϕ the closer the agent's credence is to the truth, while if ϕ is false, then the lower the agent's credence in ϕ the closer the agent's credence is to the truth. And an attractive thought is that just as truth is a primary good for binary belief, closeness to the truth is a primary good for credences.

The strategy of so-called *accuracy based* arguments for probabilism is to use the thought that closeness to truth is an epistemic good to argue that epistemic rationality demands that an agent's credences be probabilistically coherent. There are number of arguments of this form that have recently been proposed. (See e.g., Joyce (1998), Joyce (2009), Leitgeb and Pettigrew (2010a), Leitgeb and Pettigrew (2010b).) I'll outline a simple version this type of argument.

Let Σ be a set of propositions. Let n be the cardinality of Σ . We can represent a possible credal state of an agent as a vector in the vector space \mathbb{R}^n . Given a bijection B from vector coordinate positions to Σ , a real number n at position i represents a credence of n in $B(i) \in \Sigma$. There will be a subset of the set of vectors, P , containing all and only the vectors representing the probabilistically coherent credences. There will be a subset of P , W , that will do double-duty representing both possible credences and possible worlds. These will be vectors such that each coordinate is either a 1 or a 0.²⁵ The set P will be the convex hull of W , i.e., the set of convex combinations of members of W .

Given this abstract geometric representation of credal states and possible worlds, we can then use measures on such a space to make precise the idea of the *distance* of a credal state from the truth at a possible world. Now there are many possible measures that can be defined on \mathbb{R}^n . But one very natural measure is the Euclidean distance between two vectors $a = \langle a_1, a_2, \dots, a_n \rangle$, and $b = \langle b_1, b_2, \dots, b_n \rangle$, $\|a - b\| =_{df} \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$. If we take Euclidean distance to capture the epistemologically relevant notion of *closeness to the truth*, then we can argue as follows. Since the following is a theorem:

EUCLIDEAN DOMINANCE For every $x \notin P$ there is some $y \in P$, such that for every $w \in W$,
 $\|w - x\| > \|w - y\|$.

we then have:

ACCURACY DOMINANCE For every probabilistically incoherent credal state, c_1 , there is some probabilistically coherent credal state, c_2 , that is such that for every possible world w , c_2 is more accurate, i.e., closer to the truth, at w , than c_1 .

What ACCURACY DOMINANCE tells us is that if one fails to be probabilistically coherent, there will always be some way of being probabilistically coherent that is *guaranteed* to make one closer to the truth. Given that truth is a primary epistemic good, it is plausible that one ought to try to be as close to the truth as possible. It follows, then, from ACCURACY DOMINANCE that one

²⁵Although not every assignment of 1 or 0 need count as a possible world depending on which propositions are in Σ .

ought not be probabilistically incoherent, i.e., that one ought to be probabilistically coherent. And, since probabilistic coherence requires that one satisfy the constraints imposed by CREDENCE₁ - CREDENCE₂, it follows that these principles provide genuine constraints on epistemic rationality.

Now this argument makes the following crucial assumptions:

- (a) That epistemic rationality demands that one ought not be in a state that is accuracy dominated.
- (b) That accuracy, i.e., closeness to the truth, is to be measured by Euclidean distance.

In justification of (a), we noted that closeness to the truth is plausibly a primary epistemic good. But, of course, there might be other epistemic goods, and these might be in conflict with being close to the truth. If this were the case, it may be that all that we can conclude is that we have a *prima facie* epistemic obligation to not be probabilistically incoherent, but that this obligation may be overridden in certain cases. In response to this worry the defender of probabilism may respond that truth is an *overriding* epistemic good. While there may be other epistemic goods, these are, in all cases, trumped by being non-accuracy dominated. How plausible this response is is a difficult question. And it isn't one that I'll try to settle now. It does, however, seem to me to be at least *prima facie* plausible. Rejecting (a), then, does not provide an obviously attractive route to justifying the rejection of CREDENCE₁ or CREDENCE₂.

In justification of (b), we offered little more than the simplicity and naturalness of the Euclidean measure of distance within our abstract mathematical framework. More sophisticated justifications have been offered. See, e.g., Horwich (1982) and Maher (1993). However, even amongst probabilists, there is disagreement about whether we should think of this measure as capturing the epistemologically relevant notion of closeness to the truth. For dissent on this point see Joyce (1998), Joyce (2009), and Gibbard (2008). Luckily for those sympathetic to probabilism, it can be shown that the Euclidean measure is not particularly special in justifying ACCURACY DOMINANCE. There is a wide class of measures which are sufficient to justify this principle. Moreover there are classes of measures with this property that all share certain features that it is plausible to think any measure capturing the intuitive notion of closeness to the truth would have. See Joyce (1998) and Joyce (2009) for in depth discussion. It is well beyond our present scope to discuss these alternative measures in any detail. Let it suffice, then, to note that rejecting CREDENCE₁ or CREDENCE₂ by rejecting (b), would also demand that one rule out numerous other possible measures on accuracy.

Rejecting the Set-up

In generating the paradox of credence (and the paradox of belief) we assumed that it is at least possible for there to be an antecedently rational agent who has *perfect* access to its own mental states (at least concerning certain propositions). One way to block the paradox would be to deny this assumption. There are two ways that this could work. Either one could deny that it is possible for an agent to have such access to its own mental states, or one could hold that such access would itself involve some form of irrationality.

Now this type of move wouldn't ultimately resolve matters if it turned out that we could easily reconstruct the paradox using other perhaps weaker assumptions about Alpha's powers of introspection. Now it turns out that there are other assumptions that we could make about Alpha's higher-order credences that would suffice to generate the paradox. Interestingly, however, we can show that some fairly natural modest weakenings to our introspective assumptions *are* sufficient to undermine the paradox.

Instead of assuming that Alpha is always certain of its exact credence in $T(\ulcorner\beta'\urcorner)$, let's instead assume that for any credence r that Alpha may have in $T(\ulcorner\beta'\urcorner)$, Alpha is certain that its credence is within some small range around r . More precisely, instead of (9), we'll assume:

$$(13) \quad \forall x \in \mathbb{R}, Cr(T(\ulcorner\beta'\urcorner)) = x \rightarrow Cr(Cr(T(\ulcorner\beta'\urcorner)) \in [a, b]) = 1$$

where $a = x - 0.05$, if $x - 0.05 > 0$, otherwise $a = 0$; and $b = x + 0.05$, if $x + 0.05 < 1$, otherwise $b = 1$.

Further we'll assume that conditional on its credence being within some range $[a, b]$, Alpha thinks that the likelihood of its credence being within some sub-range is proportional to the size of the subrange. More specifically, we'll assume that:

$$(14) \quad Cr(Cr(T(\ulcorner\beta'\urcorner)) > 0.9 | Cr(T(\ulcorner\beta'\urcorner)) \in [a, b]) = \frac{l-u}{a-b}$$

where l is the g.l.b. and u the l.u.b. of $(0.9, 1] \cap [a, b]$.

When we assumed that (9) held, we were able to show that no matter what credence Alpha has in $T(\ulcorner\beta'\urcorner)$ it must have a different credence in $\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9$ (at least if it meets certain rational requirements). Given our new assumptions (13) and (14), however, it is easy to show that Alpha can, indeed, rationally have the same credence in $T(\ulcorner\beta'\urcorner)$ and $\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9$. That is, we can show that there is a way for Alpha to satisfy the requirements imposed by CREDENCE₁ without thereby violating a requirement imposed by CREDENCE₂.

Let $Cr(T(\ulcorner\beta'\urcorner)) = x$. We can consider three cases:

- (i) $x \leq 0.85$
- (ii) $x > 0.95$
- (iii) $0.85 < x \leq 0.95$

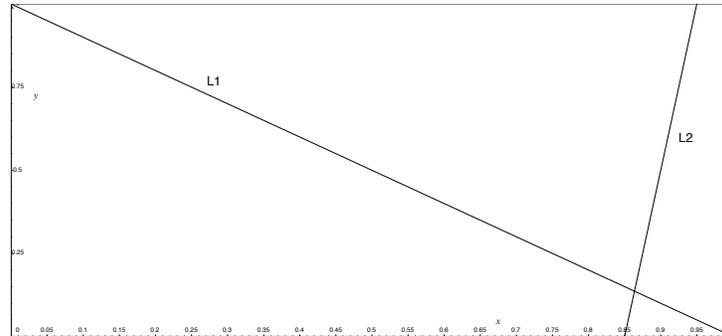
In case (i), it is clear that $Cr(Cr(T(\ulcorner\beta'\urcorner)) > 0.9) = 0$, and so assuming that Alpha meets the requirement imposed by CREDENCE₂, we have: $Cr(\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9) = 1$, and so $Cr(\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9) \neq x$.

In case (ii) it is clear that $Cr(Cr(T(\ulcorner\beta'\urcorner)) > 0.9) = 1$, and so assuming that Alpha meets the requirement imposed by CREDENCE₂, we have: $Cr(\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9) = 0$, and so $Cr(\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9) \neq x$.

How about in case (iii)? In this case we have that $Cr(Cr(T(\ulcorner\beta'\urcorner)) > 0.9) = (x - 0.85) \times 10$. It's easy to see that there is a value x such that $(x - 0.85) \times 10 = 1 - x$. This is $\frac{9.5}{11}$, or (approx.) 0.86. What this equality tells us is that if Alpha's credence in $T(\ulcorner\beta'\urcorner)$ is $\frac{9.5}{11}$, then Alpha can have

equal credence in $T(\Gamma\beta'\neg)$ and in $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$ without violating CREDENCE₂. Alpha can in this way meet the requirements imposed by CREDENCE₁ - CREDENCE₂.

This is just one sample case in which we take Alpha to have pretty good though not perfect access to its own credence in $T(\Gamma\beta'\neg)$. There is, however, a general lesson to be drawn from this case. To see this consider the following graph:



Here the x values represent Alpha's credence in $T(\Gamma\beta'\neg)$. L1 is the function $1 - x$. L2 represents Alpha's credence in $Cr(T(\Gamma\beta'\neg)) > 0.9$ as a function of Alpha's credence in $T(\Gamma\beta'\neg)$. Given assumptions (13) and (14), it follows that Alpha's credence in $Cr(T(\Gamma\beta'\neg)) > 0.9$ increases as a continuous function of Alpha's credence in $T(\Gamma\beta'\neg)$. The point at which L2 intersects L1 is the value at which Alpha's credence in $Cr(T(\Gamma\beta'\neg)) > 0.9$ is exactly $1 - x$, i.e., $1 -$ Alpha's credence in $T(\Gamma\beta'\neg)$. At this value Alpha's credence in $T(\Gamma\beta'\neg)$ and Alpha's credence in $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$ can be equal without there being a violation of CREDENCE₂.

This graph makes clear two important facts:

First, any *continuous* function of x will be guaranteed to intersect the function $1 - x$. This corresponds to the fact that one can't draw an unbroken line from 0 to 1 (on the x -axis) without crossing L1. The conclusion that we can draw from this fact is that as long as Alpha's credence in $Cr(T(\Gamma\beta'\neg)) > 0.9$ is a continuous function of Alpha's credence in $T(\Gamma\beta'\neg)$, then there will be some value at which Alpha can have the same credence in $T(\Gamma\beta'\neg)$ and $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$ without violating CREDENCE₂.

Second, there are many ways in which Alpha's credence in $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$ could be a *non-continuous* function of Alpha's credence in $T(\Gamma\beta'\neg)$, which would prevent Alpha from satisfying the demands imposed by CREDENCE₁ - CREDENCE₂. Indeed, every non-continuous function of x which does not intersect $1 - x$ will suffice. One of these functions is that which Alpha has in virtue of its assumed perfect introspective access to its credence in $T(\Gamma\beta'\neg)$. For all point less than or equal to 0.9 this function has value 0, while for all points greater than 0.9 it has value 1. There are, however, countless other ways in which Alpha's credence in $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$ could evolve given its credence in $T(\Gamma\beta'\neg)$ that would also create trouble for CREDENCE₁ and CREDENCE₂.

If one is inclined to locate the problem posed by the paradox of credence in the possible existence of a rational agent such as Alpha, one will not only need to deny the possibility of an

antecedently rational agent such as Alpha, but also the possibility of any antecedently rational agent whose credence in $\neg Cr(T(\ulcorner\beta'\urcorner)) > 0.9$ is described as one of the members of the problematic class of non-continuous functions of $T(\ulcorner\beta'\urcorner)$.

Now as a matter of psychology it might indeed seem plausible that an actual agent's credence in their credence in a proposition ϕ being above some threshold will be a continuous function of the agent's credence in ϕ . Nonetheless, it is hard to see why we should assume that an agent simply *couldn't* have the sort of access we have assumed Alpha does to its own mental state. More generally, it is hard to see why we should hold, as a matter of the metaphysics of credal states, that no agent could have high-order credences that were problematic non-continuous functions of their first-order credences. Now there might be some principled story to tell here. But I can't tell what that story would look like. And without such a story, simply ruling out all of the problematic functions as metaphysically impossible strikes me as unilluminating and implausible.

Similarly, if we are to rule out every such function as being irrational, we would need a principled story. Now here I can at least see in rough outline how one might try to rule out certain non-continuous functions as irrational. It has been thought that rationality demands that an agent be somewhat sensitive to its own mental state.²⁶ If this were the case then we could certainly rule out a number of non-continuous functions as being simply too far off the mark to count as rational. However, this type of story would clearly do nothing to rule out Alpha's introspective capacities as irrational. So what we want is a story that will explain to us why an agent being perfectly sensitive to certain aspects of its own credal state should *itself* be prohibited by rationality. It is hard for me to see what a plausible story to this effect would look like.

3.4.3 Responses to the Paradox of Knowledge

In §3.3 we saw that FACTIVITY and DEDUCTION led to a contradiction. Assuming classical logic for statements about knowledge, the argument there leaves us with two options:

- Reject FACTIVITY
- Reject DEDUCTION

Now if we are forced to choose between these two, it is, I think, fairly clear that it is DEDUCTION that should go. Although it would be hard to give an *argument* for this, FACTIVITY just seems to be a more central principle about knowledge.

Now here is a surprising consequence of holding on to FACTIVITY and classical logic for the relevant propositions, and instead rejecting DEDUCTION. Given FACTIVITY and the validity of *reductio* reasoning for the relevant propositions, it would seem that everyone, *except for Alpha*, who goes through the appropriate derivation could come to know that Alpha does not know that $\ulcorner\kappa\urcorner$ is true.

What would seem to follow is that principles about knowledge conferring warrant are *invidious* in the following sense: There are possible situations in which two agents, A_1 and A_2 , located

²⁶For example, this position seems to be endorsed by Hintikka (1962) and Elga and Egan (2005).

in the same world, have the exactly the same evidence for a proposition ϕ , and A_1 knows ϕ , but A_2 does not. Just take a perfect qualitative duplicate of Alpha, located in the same possible world as Alpha, call it *Omega*. It would seem that if Omega goes through the appropriate derivation there would be no reason to deny that Omega knows that $\lceil \kappa \rceil$ is true. But Alpha, who will have gone through the same derivation, will not know ϕ . In this case Alpha and Omega's evidence would seem to be the same, and yet one knows that ϕ while the other does not.²⁷ Thus there would seem to be possible situations in which two agents located in the same world with the same evidence are such that only one knows ϕ .

Now we may allow that there are possible cases in which two agents located in different worlds may have the same evidence for a proposition ϕ , and yet only one of them knows ϕ . Since knowledge is factive this should be possible if the agents' evidence is such as to constitute knowledge given the truth of ϕ , but ϕ is true in only one the possible worlds. We may also (although this is really quite contentious) want to allow for so-called pragmatic encroachment, where facts not just about the agent's evidential situation, but also facts about the agent's practical situation have an effect on whether or not an agent knows a proposition. What we would seem to be forced to acknowledge in addition to these facts, though, is that in certain situations it can be *the mere fact that it is A_1* that prevents the agent from knowing some proposition ϕ .

It is surprising that knowledge could be invidious in this sense. Of course, this may simply be something that we have to live with, but *ceteris paribus* it would be preferable if we could allow for a tighter connection between an agent's evidential situation and its knowledge.

²⁷One might worry here that, at least according to views on which one's evidence just is one's knowledge, Alpha and Omega won't have the same evidence. For, of course, Omega knows that $\lceil \kappa \rceil$ is true, but Alpha does not. Granted. But the conclusion that there is a peculiar invidious quality to the connection between evidence and knowledge can be reformulated. Simply consider Omega's evidential base minus the fact that $\lceil \kappa \rceil$ is true. Call this E . If Omega knows that $\lceil \kappa \rceil$ is true in this case, this fact is presumably going to be grounded in facts about E , e.g., knowledge about the valid derivation of the truth of $\lceil \kappa \rceil$. So we get the following conclusion: two agent's can both have evidence E which for only one agent is sufficient to ground knowledge in a proposition ϕ , and the failure to ground knowledge in the other agent's case simply stems from the fact that it is the particular agent in question who has the evidence, and not, say from other evidence that the agent has which could undermine or defeat the support E lends to ϕ .

Chapter 4

Non-Classical Solutions

We have considered a series of attitudinal paradoxes that show that a number of *prima facie* plausible principles governing the attitudes of belief, credence and knowledge are *classically* inconsistent. We've considered various ways in which these paradoxes can be resolved that don't involve appealing to non-classical logical resources. In each case, I've argued that the solution brings with it theoretical costs—in certain cases fairly significant costs.

Now, it may be that classical logic is correct, and so these costs must be incurred one way or another. However, there are good reasons to take seriously various non-classical logics. As we've already seen, the liar paradox gives us very good (perhaps our best) reason to invest at least some positive credence in certain non-classical logics. In chapter 2, we considered a number of non-classical logics that could be motivated by appealing to this paradox. In this chapter, we'll see how these non-classical logics can be used to provide elegant solutions to the paradoxes developed in chapter 3. Appealing to such resources will allow us to hold on to all of our plausible principles governing doxastic, credal and epistemic states.

This, of course, doesn't show us that we must reject classical logic. It does, however, give us some further reason for thinking that some non-classical logic may be correct. For our purposes, however, the more important conclusions to be drawn are conditional in character. What I think these arguments show is that *if* one endorses one of these non-classical logics, then one should appeal to such logics to solve these paradoxes. These conditionals will turn out to have important consequences later on.

4.1 Non-classical Solutions to the Paradox of Belief

4.1.1 The Structural Identity of Certain Doxastic and Semantic Paradoxes

To see how appeal to non-classical resources can be of use in solving the paradox developed in §3.1, it will help to first take a brief look at a related semantic paradox. Let $\ulcorner \eta \urcorner$ name the following sentence: $\neg \Box T(\ulcorner \eta \urcorner)$. On the assumption that the logic governing \Box is S5 we can derive a contradiction from this sentence as follows:

1	$T(\ulcorner \eta \urcorner) \leftrightarrow \neg \Box T(\ulcorner \eta \urcorner)$	T-schema
2	$\Box(\Box T(\ulcorner \eta \urcorner) \rightarrow \neg T(\ulcorner \eta \urcorner))$	1, Nec.
3	$\Box(\neg \Box T(\ulcorner \eta \urcorner) \rightarrow T(\ulcorner \eta \urcorner))$	1, Nec.
4	$\Box \Box T(\ulcorner \eta \urcorner) \rightarrow \Box \neg T(\ulcorner \eta \urcorner)$	2, Ax.K
5	$\Box \neg \Box T(\ulcorner \eta \urcorner) \rightarrow \Box T(\ulcorner \eta \urcorner)$	3, Ax.K
6	$\Box T(\ulcorner \eta \urcorner) \rightarrow \Box \Box T(\ulcorner \eta \urcorner)$	Ax.4
7	$\neg \Box T(\ulcorner \eta \urcorner) \rightarrow \Box \neg \Box T(\ulcorner \eta \urcorner)$	Ax.5
8	$\Box \neg T(\ulcorner \eta \urcorner) \rightarrow \neg \Box T(\ulcorner \eta \urcorner)$	S5 theorem
9	$\Box T(\ulcorner \eta \urcorner) \vee \neg \Box T(\ulcorner \eta \urcorner)$	Classical Theorem
10	$\Box T(\ulcorner \eta \urcorner)$	
11	$\Box \Box T(\ulcorner \eta \urcorner)$	6, 10
12	$\Box \neg T(\ulcorner \eta \urcorner)$	4, 11
13	$\neg \Box T(\ulcorner \eta \urcorner)$	8, 12
14	\perp	10, 13
15	$\neg \Box T(\ulcorner \eta \urcorner)$	
16	$\Box \neg \Box T(\ulcorner \eta \urcorner)$	7, 15
17	$\Box T(\ulcorner \eta \urcorner)$	5, 16
18	\perp	15, 17
19	\perp	9, 10 - 14, 15 - 18

For ease of reference later, I'll refer to the steps in this derivation as $(1\eta) - (19\eta)$. (1η) is an instance of the T-schema. (2η) and (3η) follow from (1η) on the assumption that the logic governing \Box is a normal modal logic, and so obeys the rule of necessitation. (4η) follows from (2η) , and (5η) from (3η) , on the assumption that the logic for \Box is a normal modal logic and so obeys axiom K: $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$. (6η) holds if the logic governing \Box obeys axiom 4: $\Box\phi \rightarrow \Box\Box\phi$. (7η) holds if the logic governing \Box obeys axiom 5: $\neg\Box\phi \rightarrow \Box\neg\Box\phi$. (8η) holds given that axiom T: $\Box\phi \rightarrow \phi$ holds. (9η) is a theorem of classical logic.

If we take \Box to express metaphysical necessity then it is very plausible that S5 is the correct logic for the operator, and so all of the above modal axioms hold. Given these principles we can derive a contradiction on the assumption $\Box T(\ulcorner \eta \urcorner)$ and on the assumption $\neg \Box T(\ulcorner \eta \urcorner)$ using simply *modus ponens*. A contradiction can then be derived outright from (9η) by proof-by-cases reasoning.

The paradox developed in §3.1 can be represented in the form of a derivation that parallels the modal liar paradox:

1	$T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$	T-schema
2	$B_\alpha(B_\alpha T(\ulcorner\beta\urcorner) \rightarrow \neg T(\ulcorner\beta\urcorner))$	Assumption
3	$B_\alpha(\neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow T(\ulcorner\beta\urcorner))$	Assumption
4	$B_\alpha B_\alpha T(\ulcorner\beta\urcorner) \rightarrow B_\alpha \neg T(\ulcorner\beta\urcorner)$	2, EVIDENCE
5	$B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow B_\alpha T(\ulcorner\beta\urcorner)$	3, EVIDENCE
6	$B_\alpha T(\ulcorner\beta\urcorner) \rightarrow B_\alpha B_\alpha T(\ulcorner\beta\urcorner)$	Assumption
7	$\neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)$	Assumption
8	$B_\alpha \neg T(\ulcorner\beta\urcorner) \rightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$	CONSISTENCY
9	$B_\alpha T(\ulcorner\beta\urcorner) \vee \neg B_\alpha T(\ulcorner\beta\urcorner)$	Classical Theorem
10	$B_\alpha T(\ulcorner\beta\urcorner)$	
11	$B_\alpha B_\alpha T(\ulcorner\beta\urcorner)$	6, 10
12	$B_\alpha \neg T(\ulcorner\beta\urcorner)$	4, 11
13	$\neg B_\alpha T(\ulcorner\beta\urcorner)$	8, 12
14	\perp	10, 13
15	$\neg B_\alpha T(\ulcorner\beta\urcorner)$	
16	$B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)$	7, 15
17	$B_\alpha T(\ulcorner\beta\urcorner)$	5, 16
18	\perp	15, 17
19	\perp	9, 10 - 14, 15 - 18

Formally this derivation is almost identical to the first; where they differ are in the justifications of certain steps.

For ease of reference later, I'll refer to the steps in this derivation as $(1\beta) - (19\beta)$. As above, (1β) is an instance of the T-schema. (2β) and (3β) correspond to our assumption that the agent believes the theorem expressed at (1β) . Given the set-up of the case, (4β) is a consequence of the assumption that Alpha meets the requirements imposed by EVIDENCE. Why? Because as we set-up the case, it follows from the evidential status of the belief codified in (2β) that if Alpha believes that it believes that $\ulcorner\beta\urcorner$ is true then Alpha will have evidence that makes it certain that $\ulcorner\beta\urcorner$ is not true. So, assuming that Alpha meets the requirements imposed by EVIDENCE, it follows that if Alpha believes that it believes that $\ulcorner\beta\urcorner$ is true then it will believe that $\ulcorner\beta\urcorner$ is not true. This is (4β) .

A similar story explains how (5 β) follows from the assumption that Alpha meets the requirements imposed by EVIDENCE. (6 β) and (7 β) are assumptions that we made about Alpha. (8 β) corresponds to our assumption that the agent meets the normative condition specified in CONSISTENCY. From these we can derive a contradiction, using *modus ponens*, on the assumption $B_\alpha T(\ulcorner\beta\urcorner)$ and on the assumption $\neg B_\alpha T(\ulcorner\beta\urcorner)$. Given the assumption that the classical validity $B_\alpha T(\ulcorner\beta\urcorner) \vee \neg B_\alpha T(\ulcorner\beta\urcorner)$ obtains, a contradiction can be derived outright by proof-by-cases reasoning.

4.1.2 The Paracomplete Solution

According to paracomplete approaches to the semantic paradoxes, excluded-middle is not valid for paradox inducing sentences. The derivation of a contradiction from $\ulcorner\eta\urcorner$, then, is blocked at (9 η). In chapter 2 we saw how to construct paracomplete models for a language \mathcal{L}^+ containing a truth predicate T , a conditional \rightarrow , and a modal operator \Box . It is easy enough to show that there are VF^+ models in which the accessibility relation is an equivalence relation, (1 η)-(8 η) all have semantic value 1 at every point, but the instance of excluded-middle for $\Box T(\ulcorner\eta\urcorner)$ has value 1/2 at each point. The existence of such models, then, assures us that a proponent of KFS^+ can indeed block the route to a contradiction from (1 η)-(8 η) by rejecting excluded-middle for $\Box T(\ulcorner\eta\urcorner)$. And, although we cannot say that $\ulcorner\eta\urcorner$ is *not* necessarily true, we can say that it is neither determinate that it is necessarily true, nor determinate that it is not necessarily true; that is, we can say that it is indeterminate whether $\ulcorner\eta\urcorner$ is necessarily true. $I\Box T(\ulcorner\eta\urcorner)$ is valid in such models.

Since the two derivations above proceed in parallel, the strategy for blocking the contradiction in the former case will work equally well in the latter. Just as we can block the derivation of a contradiction in the first case by giving up excluded-middle for $\ulcorner\eta\urcorner$, so too can we block the derivation of a contradiction in the second case by giving up excluded-middle for $\ulcorner\beta\urcorner$.

Allowing that Alpha's doxastic state may be such that excluded-middle fails for the claim that it believes that $\ulcorner\beta\urcorner$ is true, allows us to defuse the argument given in §3.1 that CONSISTENCY, EVIDENCE and POSSIBILITY are incompatible. It is true that, given the assumption that Alpha meets the requirements imposed by both CONSISTENCY and EVIDENCE, we can derive a contradiction on the assumption that Alpha believes that $\ulcorner\beta\urcorner$ is true and on the assumption that Alpha doesn't believe that $\ulcorner\beta\urcorner$ is true. This was taken to show that CONSISTENCY and EVIDENCE were in conflict with POSSIBILITY. Crucially, however, this requires that we be able to help ourselves to the claim that excluded-middle holds for the proposition that Alpha believes that $\ulcorner\beta\urcorner$ is true. If, then, excluded-middle fails for this proposition, it will not follow from the fact that we can derive a contradiction from the assumption that Alpha meets the requirements imposed by CONSISTENCY and EVIDENCE, *given* that Alpha believes that $\ulcorner\beta\urcorner$ is true, and *given* that Alpha does not believe that $\ulcorner\beta\urcorner$ is true, that we can derive a contradiction from this assumption outright. We need not infer from the possibility of an agent such as Alpha that CONSISTENCY and EVIDENCE entail a violation of POSSIBILITY.

Of course, given the explicit appeal to excluded-middle in the development of the normative paradox, it is obvious that a paracomplete theorist can block the paradox as stated. But importantly the paracomplete theorist can do more than this. What the paradox developed in §3.1 purported to show is that there is a possible case in which an antecedently rational agent cannot meet the

requirements imposed by CONSISTENCY and EVIDENCE. Using paracomplete resources, however, we can assure ourselves that an agent can in fact satisfy the stipulations made about Alpha and also meet all the requirements imposed by CONSISTENCY and EVIDENCE.

One way to think about the paradox developed in §3.1 is as follows. Given that Alpha knows that $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$, it follows from the transparency assumptions, (3) and (4), that, whether or not Alpha believes that $\ulcorner\beta\urcorner$ is true, EVIDENCE will impose a requirement that some of Alpha's beliefs be closed under logical consequence. But meeting this local closure requirement is either impossible (in the case in which Alpha does not believe that $\ulcorner\beta\urcorner$ is true) or leads to a violation of the requirement that the agent satisfy CONSISTENCY (in the case in which Alpha does believe that $\ulcorner\beta\urcorner$ is true).

Viewing the problem in terms of a local closure requirement lets us see more clearly how an appeal to paracomplete resources can help resolve the problem. For, using the paracomplete model-theory developed to deal with the modal liar sentence to interpret Alpha's belief operator, we can provide a model in which the following all hold:

$$(3) B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow B_\alpha B_\alpha T(\ulcorner\beta\urcorner)$$

$$(4) \neg B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)$$

$$(15) B_\alpha (T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner))$$

$$(16) \text{Every instance of the following schema is satisfied: } B_\alpha \phi \rightarrow \neg B_\alpha \neg \phi$$

$$(17) \text{Alpha's beliefs are closed under logical consequence}$$

In any such model, excluded-middle must fail for the claim that Alpha believes that $\ulcorner\beta\urcorner$ is true. Indeed, in the simplest such models it will be indeterminate whether Alpha believes that $\ulcorner\beta\urcorner$ is true.

We can show, then, that if excluded-middle fails for the claim that Alpha believes that $\ulcorner\beta\urcorner$ is true, Alpha can satisfy the transparency assumptions, know that $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$, and yet not be forced into violating CONSISTENCY in order to meet EVIDENCE. If, then, we allow for this failure of excluded-middle, we can hold on to what are some seemingly quite basic normative principles.

Moreover it should be clear that this is really the only way for a paracomplete theorist to hold on to CONSISTENCY, EVIDENCE and POSSIBILITY (assuming that such a theorist accepts the possibility of an antecedently rational agent such as Alpha). For, on the assumption that Alpha meets the requirements imposed by CONSISTENCY and EVIDENCE, a paracomplete theorist will accept (1 β)-(8 β). In addition, this theorist will accept reasoning by modus ponens and disjunction elimination. But, as the derivation makes clear, given these commitments, the only way to avoid a contradiction is to allow that excluded-middle fails for $\ulcorner\beta\urcorner$.

Let M be a VF^+ model for which R_m is an equivalence relation, i.e., a transitive, reflexive and symmetric relation. Let B_α be treated in the model in the manner specified for \Box in §2.2. For all g

and δ the following hold:

Proof of (3)

$$\llbracket B_\alpha T(\Gamma\beta^\top) \leftrightarrow B_\alpha B_\alpha T(\Gamma\beta^\top) \rrbracket^{g,\delta} = 1$$

For each stage α in the Field construction, at the resulting Kripke model M_α , $\llbracket B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta} = \llbracket B_\alpha B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta}$. So at each stage $\alpha > 0$, $\llbracket B_\alpha T(\Gamma\beta^\top) \leftrightarrow B_\alpha B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta} = 1$.

Proof of (4)

$$\llbracket \neg B_\alpha T(\Gamma\beta^\top) \leftrightarrow B_\alpha \neg B_\alpha T(\Gamma\beta^\top) \rrbracket^{g,\delta} = 1$$

As above, for each stage α in the construction, $\llbracket \neg B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta} = \llbracket B_\alpha \neg B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta}$. So at each stage $\alpha > 0$, $\llbracket \neg B_\alpha T(\Gamma\beta^\top) \leftrightarrow B_\alpha \neg B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta} = 1$.

Proof of (15)

$$\llbracket B_\alpha (T(\Gamma\beta^\top) \leftrightarrow \neg B_\alpha T(\Gamma\beta^\top)) \rrbracket^{g,\delta} = 1$$

For each stage $\alpha > 0$ in the construction, and each point δ' , $\llbracket T(\Gamma\beta^\top) \leftrightarrow \neg B_\alpha T(\Gamma\beta^\top) \rrbracket_{M_\alpha}^{g,\delta'} = 1$. It follows that for every stage $\alpha > 0$, $\llbracket B_\alpha (T(\Gamma\beta^\top) \leftrightarrow \neg B_\alpha T(\Gamma\beta^\top)) \rrbracket_{M_\alpha}^{g,\delta} = 1$.

Proof of (16)

$$\llbracket B_\alpha \phi \rightarrow \neg B_\alpha \neg \phi \rrbracket^{g,\delta} = 1$$

Assume $\llbracket B_\alpha \phi \rrbracket_{M_\alpha}^{g,\delta} = 1$. It follows that for every point δ' , $\llbracket \phi \rrbracket_{M_\alpha}^{g,\delta'} = 1$, and so $\llbracket \neg \phi \rrbracket_{M_\alpha}^{g,\delta'} = 0$. It follows that $\llbracket B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta} = 0$ and so $\llbracket \neg B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta} = 1$.

Assume $\llbracket B_\alpha \phi \rrbracket_{M_\alpha}^{g,\delta} = 1/2$. It follows that for no point δ' is it the case that $\llbracket \phi \rrbracket_{M_\alpha}^{g,\delta'} = 0$ and for some point δ' it is the case that $\llbracket \phi \rrbracket_{M_\alpha}^{g,\delta'} = 1/2$. From this it follows that every point δ' is such that either $\llbracket \neg \phi \rrbracket_{M_\alpha}^{g,\delta'} = 0$ or $\llbracket \neg \phi \rrbracket_{M_\alpha}^{g,\delta'} = 1/2$. Thus, $\llbracket B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta} \leq 1/2$, and so $\llbracket \neg B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta} \geq 1/2$.

For every stage α , then, $\llbracket B_\alpha \phi \rrbracket_{M_\alpha}^{g,\delta} \leq \llbracket \neg B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta}$. And so for every stage $\alpha > 0$ $\llbracket B_\alpha \phi \rightarrow \neg B_\alpha \neg \phi \rrbracket_{M_\alpha}^{g,\delta} = 1$.

Proof of (17)

Let Γ be a set of sentences. Let $B_\alpha \Gamma$ stand for the set of sentences that result by appending B_α to each member of Γ . We'll show that $\Gamma \models \psi \Rightarrow B_\alpha \Gamma \models B_\alpha \psi$

Assume that $\Gamma \models \psi$. Take an arbitrary model M in the class of VF^+ models and assume that in this model $\llbracket B_\alpha \gamma \rrbracket_M^\delta = 1$ for every $B_\alpha \gamma \in B_\alpha \Gamma$. We'll argue that $\llbracket B_\alpha \psi \rrbracket_M^\delta = 1$. At every point δ' such that $\delta R \delta'$, $\llbracket \gamma \rrbracket_M^\delta = 1$, for every $\gamma \in \Gamma$. Given that $\Gamma \models \psi$ it follows that at every such δ' $\llbracket \psi \rrbracket_M^{\delta'} = 1$. And so $\llbracket B_\alpha \psi \rrbracket_M^\delta = 1$. Thus $B_\alpha \Gamma \models B_\alpha \psi$.

4.1.3 The Supervaluationist Solution

According to supervaluationist approaches to the semantic paradoxes excluded-middle *is* valid for paradox inducing sentences. However, proof-by-cases reasoning is not unrestrictedly valid. The derivation of a contradiction from $\ulcorner \eta \urcorner$, then, is blocked not at (9 η), but at the final step where it is inferred that (19 η) holds by appeal to (9 η), together with the subproofs (10 η) – (14 η) and (15 η) – (18 η). The supervaluationist can grant that (1 η) – (9 η) hold and that given these premisses (14 η) follows from (10 η) and (18 η) follows from (15 η). But crucially she will deny that this means that (19 η) holds.

It is easy enough to verify this by appeal to our model theory. In any SV^+ model in which the accessibility relation is an equivalence relation (1 η) – (9 η) will have semantic value 1. And it's easy to show that in any SV^+ in which (1 η) – (9 η) have semantic value 1, if (10 η) has semantic value 1, then so too will (14 η), and if (15 η) has semantic value 1 then so too will (18 η). Given the existence of such models, then, the proponent of SV^+ can endorse every part of our derivation except for the conclusion. The key fact is that the following meta-inference:

$$\Gamma, \phi \models \psi \wedge \Gamma, \neg\phi \models \psi \Rightarrow \Gamma \models \psi$$

fails if validity is understood in the way suggested by the proponent of SV^+ . In the simplest SV^+ models in which this holds, it will turn out that $I \Box T(\ulcorner \eta \urcorner)$ has value 1.

Again, given the formal identity of the the two paradoxes, the same strategy for blocking the modal liar paradox can in principle be extended to block the epistemic paradox. Just as we can block the derivation of a contradiction from $\ulcorner \eta \urcorner$ by rejecting the appeal to proof-by-cases reasoning, so too can we block the derivation of a contradiction from $\ulcorner \beta \urcorner$.

Of course, given the explicit appeal to proof-by-cases reasoning in our statement of the paradox in §3.1 this is not surprising. But it is important again to note that the supervaluationist, like the paracomplete theorist, can do more than simply block the paradox as stated. Using the supervaluationist model theory we can assure ourselves that an agent can indeed meet all the requirements that we imposed on Alpha, and in addition meet all of the normative demands imposed by CONSISTENCY and EVIDENCE.

In the paracomplete case, we were able to show that there are VF^+ models in which (3), (4) and (15) - (17) all hold. The same can be shown using supervaluationist models. Indeed the proofs offered for these claims all go through if we simply switch from VF^+ models to SV^+ models. Any such model will serve as a counter-example to the validity of proof-by-cases reasoning. And, in the simplest such models it will be indeterminate whether Alpha believes that $\ulcorner \beta \urcorner$ is true.

Given the existence of such models, we can be assured that if we allow that Alpha's doxastic state may give rise to the same failures of classical logic that the supervaluationist thinks are occasioned by paradoxical propositions such as that expressed by $\ulcorner \eta \urcorner$, then Alpha can satisfy all of the requirements imposed by CONSISTENCY and EVIDENCE, despite its knowledge of certain theorems and its introspective powers. Once again, then, we need not, if we are proponents of such a non-classical theory, infer from our earlier normative paradox that CONSISTENCY and EVIDENCE are incompatible with POSSIBILITY.

Moreover, this is the only way for a supervaluationist to hold on to CONSISTENCY, EVIDENCE and POSSIBILITY (again, assuming that such a theorist accepts the possibility of an antecedently rational agent such as Alpha). For, on the assumption that Alpha meets CONSISTENCY and EVIDENCE, a supervaluationist will accept (1 β)-(9 β). Moreover the supervaluationist will accept the inferences appealed to in subproofs 1 and 2. But, as the derivation makes clear, given these commitments, the only way to avoid a contradiction is to reject proof-by-cases reasoning for the claim that Alpha believes that $\ulcorner\beta\urcorner$ is true.

4.1.4 The Paraconsistent Solution

A paraconsistent theorist can also provide a solution to our modal liar paradox. Seeing how this works, however, requires some distinctions that were not important for the paracomplete and paraconsistent theories. In the development of the modal liar paradox we used \perp to stand for some arbitrary absurdity. The key fact about \perp is that for any formula ϕ the following holds: $\perp \models \phi$, i.e., \perp entails everything. Now, according to classical, paracomplete and supervaluationist accounts of validity (at least the ones that we're interested in) any contradiction can play the role of \perp . That is for any contradiction $\psi \wedge \neg\psi$, we have, according to these theories, for arbitrary ϕ :

EXPLOSION $\psi \wedge \neg\psi \models \phi$.

According to a paracomplete theorist, however, EXPLOSION is not a valid schema. While \perp and other contradictions may indeed entail everything, this is not true of every contradiction. Since \perp does entail everything, if we want to deny that a particular contradiction satisfies EXPLOSION, we must deny that from that contradiction \perp follows. In the case of our modal liar paradox, then, the paraconsistent solution is to reject the inference from (10 η) and (13 η) to (14 η) and the inference from (15 η) and (17 η) to (18 η).

It is easy to verify this by appeal to our model theory. We already noted that for VF^+ models in which the accessibility relation is an equivalence relation (1 η) – (8 η) all have semantic value 1, and (9 η) has semantic value 1/2. In these models (10 η), (13 η) (and so (15 η) and (17 η)) will also have semantic value 1/2. In any such model, however, \perp will receive semantic value 0. The proponent of LP^+ takes it that validity demands that the conclusion always have at least as great a semantic value as the lowest value of the premisses. What such a model shows, then, is that by paraconsistent lights \perp doesn't follow from (10 η) and (13 η), (or equivalently from (15 η) and (17 η)).

In the appropriate models in which (1 η) – (9 η) all receive a designated value (i.e., values 1 or 1/2), $\Box T(\ulcorner\eta\urcorner) \wedge \neg\Box T(\ulcorner\eta\urcorner)$ will receive semantic value 1/2. Assuming an S5 modal logic, then, the paraconsistent theorist will hold that $\Box T(\ulcorner\eta\urcorner) \wedge \neg\Box T(\ulcorner\eta\urcorner)$ is one of the contradictions for which EXPLOSION fails.

Now, again, given the formal identity of the modal liar paradox and our normative paradox, it is at least in principle open to the paraconsistent theorist to resolve the latter in the same manner as the former. Indeed, using VF^+ models, the paraconsistent theorist can show, in exactly the same manner as the paracomplete theorist, that at least in principle an agent can satisfy (3), (4), and (15)

- (17). That is, the paracomplete theorist can show that, at least in principle, an agent could satisfy all of the stipulations that were made about Alpha without thereby being forced to violate either CONSISTENCY or EVIDENCE. The difference between the paracomplete and the paraconsistent theorist is that the latter will hold that for this to be the case Alpha must both believe that $\ulcorner\beta\urcorner$ is true and not believe that $\ulcorner\beta\urcorner$ is true.

It might seem, however, that the paraconsistent theorist can offer a simpler resolution of the paradox developed in §3.1. There, it was argued that an agent such as Alpha could not meet the normative demands imposed by CONSISTENCY and EVIDENCE. Now CONSISTENCY, one might think, is a plausible principle for theories that *don't* accept contradictions. But it might seem that if one accepts that certain contradictions are true, then one shouldn't accept that CONSISTENCY provides a general normative constraint on beliefs. Thus, it might seem that while the epistemic paradox developed in §3.1 is pressing for classical, paracomplete and supervaluationist theorists, it has little force for a paraconsistent theorist.

This line of thought is tempting. But things aren't as simple as they might appear.

First, I think that the question of whether or not a paraconsistent theorist should accept CONSISTENCY is actually quite subtle. As we've already seen, rejecting CONSISTENCY isn't *required* in order for the paraconsistent theorist to resolve the epistemic paradox. And I'll argue later that there are good reasons for such a theorist to accept that CONSISTENCY provides a general normative constraint on beliefs.

Second, and more importantly, the question of whether or not a paraconsistent theorist should accept CONSISTENCY is orthogonal to my present purposes. The point that I want to make right now is that in order for the paraconsistent theorist to resolve the normative paradox in §3.1, she must hold that Alpha, if it is rational, will both believe and not believe that $\ulcorner\beta\urcorner$ is true. And this we can argue for whether or not we accept CONSISTENCY.

We've already seen how Alpha can meet the normative requirements imposed by CONSISTENCY and EVIDENCE, if Alpha both believes and doesn't believe that $\ulcorner\beta\urcorner$ is true. And it's easy to see that this is the only way for Alpha to meet the requirements imposed by CONSISTENCY and EVIDENCE. For, on the assumption that Alpha meets these requirements, $(1\beta) - (8\beta)$ will hold. Moreover, the paraconsistent theorist will accept (9β) . And from $(1\beta) - (9\beta)$ there is a valid paraconsistent derivation of $B_\alpha T(\ulcorner\beta\urcorner) \wedge \neg B_\alpha T(\ulcorner\beta\urcorner)$. (Simply replace \perp in the above derivation with $B_\alpha T(\ulcorner\beta\urcorner) \wedge \neg B_\alpha T(\ulcorner\beta\urcorner)$.) We can conclude that if the paraconsistent theorist accepts CONSISTENCY, then she must accept that in order to resolve the paradox in §3.1 Alpha must be such that it both does and does not believe that $\ulcorner\beta\urcorner$ is true.

If, however, the paracomplete theorist thinks that we should resolve the paradox by simply rejecting the appropriate instances of CONSISTENCY, we can still argue that she should hold that Alpha must both believe and not believe that $\ulcorner\beta\urcorner$ is true. Call a true contradiction a *dialethia*.¹ Although a paraconsistent theorist thinks that there are dialethias and so classical logic must be abandoned, she will nonetheless hold that there are certain restricted domains in which classical logic does hold. Now even if a paraconsistent theorist doesn't accept that CONSISTENCY provides a correct *general* normative constraint on beliefs, she should still, nonetheless, hold that in those

¹This terminology is from Priest (2006b).

cases in which classical logic holds this type of principle does apply. In order, then, for such a theorist to reject a particular instance of CONSISTENCY, the theorist must hold that the propositions concerned give rise to dialethias. The only instance of CONSISTENCY that we appealed to in our paradox was the following: $O(B_\alpha \neg T(\ulcorner \beta \urcorner) \rightarrow \neg B_\alpha T(\ulcorner \beta \urcorner))$. If the paraconsistent theorist, then, is to reject this instance of consistency, she must hold that $T(\ulcorner \beta \urcorner)$ gives rise to a dialethia. That is she must hold that we have both $T(\ulcorner \beta \urcorner)$ and $\neg T(\ulcorner \beta \urcorner)$. But, since we have $T(\ulcorner \beta \urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner \beta \urcorner)$, it follows that Alpha must both believe that $\ulcorner \beta \urcorner$ is true and not believe that $\ulcorner \beta \urcorner$ is true.

Whether or not the paraconsistent theorist wants to reject certain instances of CONSISTENCY, we can still conclude that, in order to resolve the paradox developed in §3.1, a paraconsistent theorist must say that Alpha must be such that it both believes and does not believe that $\ulcorner \beta \urcorner$ is true. A paraconsistent theorist, then, can resolve the epistemic paradox; however, as with the paracomplete theorist and the supervaluationist, doing so requires that allowing that Alpha's belief state may give rise to the same types of failures of classical logic as arise for paradoxical propositions such as that expressed by the liar sentence.

4.2 Non-classical Solutions to the Paradox of Credence

4.2.1 The Paracomplete Solution

In response to the paradox of belief, we saw how one could provide a paracomplete model of the doxastic state of an agent meeting all of stipulations made about Alpha, in which the agent was also able to meet all of the requirements imposed by EVIDENCE and CONSISTENCY. What we will show here is that a paracomplete theorist can provide a similar solution to the paradox of credence.

So far, however, we have no way of providing models for a language in which facts about credence can be expressed. In order, then, to provide our paracomplete solution we need to extend our model theory.

Let us now denote by \mathcal{L} a first-order language containing (in addition to a truth-predicate, the normal first-order quantifiers and boolean connectives) the following expressions: $Cr(\cdot) \leq x$ and $Cr(\cdot) \geq x$. These expressions are such that we get a wff of \mathcal{L} if \cdot is replaced by a wff of \mathcal{L} , and x is replaced by a term of \mathcal{L} . We let $(Cr(\cdot) = x) =_{df} (Cr(\cdot) \leq x \wedge Cr(\cdot) \geq x)$. We will let \mathcal{L}^+ be the result of adding the binary connective \rightarrow to \mathcal{L} .

We will now show how to construct *VF*-type models for \mathcal{L} and \mathcal{L}^+ .

A *VF* model for \mathcal{L} will be a tuple $\langle D_m, F, \Delta_m, \mathcal{P}, \llbracket \cdot \rrbracket_m \rangle$. The only new item here is \mathcal{P} .² This is a probability measure over Δ_m . That is \mathcal{P} is a function mapping subsets of Δ_m to real numbers in $[0, 1]$, such that $\mathcal{P}(\Delta_m) = 1$, $\mathcal{P}(\emptyset) = 0$, and for any two disjoint subsets of Δ_m , Q, R , $\mathcal{P}(P \cup Q) = \mathcal{P}(P) + \mathcal{P}(Q)$.³ As in the earlier cases, the value space of our new *VF* models will be $\{0, 1/2, 1\}$. The semantic values of atomic formulas, quantified formulas and formulas with boolean

²I've also removed the accessibility relation R , since it will play no role in the characterization of the semantic value of either $Cr(\cdot) \leq x$ or $Cr(\cdot) \geq x$.

³I'll assume that every subset of Δ_m is measurable.

connectives as their operator (relative to a sequence and point of evaluation) will be determined as in our earlier VF models. To specify the class of models that we'll be interested in we simply need to specify how the semantic values of formulas of the form $Cr(\psi) \leq x$ and $Cr(\psi) \geq x$ are determined.

Let $\Sigma^{\llbracket \psi \rrbracket^g = 1}$ be the subset of points in Δ_m in which ϕ receives semantic value 1 (relative to g). Similarly for $\Sigma^{\llbracket \psi \rrbracket^g = 1/2}$ and $\Sigma^{\llbracket \psi \rrbracket^g = 0}$.

We say:

- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = Cr(\psi) \leq x$ then:
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \leq \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 0$ iff either $\llbracket x \rrbracket_m^{g,\delta} \notin \mathbb{R}$ or for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \not\leq \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1/2$, otherwise, i.e., if for some $Q, Q' \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \leq \llbracket x \rrbracket_m^{g,\delta}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q') \not\leq \llbracket x \rrbracket_m^{g,\delta}$.
- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = Cr(\psi) \geq x$ then:
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \geq \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 0$ iff either $\llbracket x \rrbracket_m^{g,\delta} \notin \mathbb{R}$ or for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \not\geq \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1/2$, otherwise, i.e., if for some $Q, Q' \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \geq \llbracket x \rrbracket_m^{g,\delta}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q') \not\geq \llbracket x \rrbracket_m^{g,\delta}$.

It is a consequence of these clauses, together with our definition of $Cr(\psi) = x$, that we have:

- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = (Cr(\psi) = x)$ then:
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) = \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 0$ iff either $\llbracket x \rrbracket_m^{g,\delta} \notin \mathbb{R}$ or for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) \neq \llbracket x \rrbracket_m^{g,\delta}$.
 - $\llbracket \phi \rrbracket_m^{g,\delta} = 1/2$, otherwise, i.e., if for some $Q, Q' \subseteq \Sigma^{\llbracket \psi \rrbracket^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q) = \llbracket x \rrbracket_m^{g,\delta}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1} \cup Q') \neq \llbracket x \rrbracket_m^{g,\delta}$.

The VF models that we will be interested in will be those that result from an underlying classical model via the standard Kripke construction. Of course, in order for such models to exist we need to be assured that MONOTONICITY still holds given the presence of $Cr(\psi) \leq x$ and $Cr(\psi) \geq x$ in our language. Showing this, however, is not hard.

We simply need to show that if α and γ are stages in our Kripke construction, such that $\alpha < \gamma$, then for every sequences g and point of evaluation δ , (i) if $\llbracket Cr(\psi) \leq x \rrbracket_\alpha^{g,\delta} = 1$ then $\llbracket Cr(\psi) \leq x \rrbracket_\gamma^{g,\delta} = 1$, and (ii) if $\llbracket Cr(\psi) \leq x \rrbracket_\alpha^{g,\delta} = 0$ then $\llbracket Cr(\psi) \leq x \rrbracket_\gamma^{g,\delta} = 0$ (and similarly for $Cr(\psi) \geq x$). Here we can sketch the argument for $Cr(\psi) \leq x$.

(i) Assume that $\llbracket Cr(\psi) \leq x \rrbracket_\alpha^{g,\delta} = 1$. We want to show $\llbracket Cr(\psi) \leq x \rrbracket_\gamma^{g,\delta} = 1$. Our assumption tells us that $\llbracket x \rrbracket_\alpha^{g,\delta} \in \mathbb{R}$, and for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket_\alpha^g = 1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket_\alpha^g = 1} \cup Q) \leq \llbracket x \rrbracket_\alpha^{g,\delta}$. As an inductive hypothesis, we can take it that MONOTONICITY holds for ψ . Thus we have (a) $\Sigma^{\llbracket \psi \rrbracket_\gamma^g = 1/2} \subseteq \Sigma^{\llbracket \psi \rrbracket_\alpha^g = 1/2}$,

and (b) $\Sigma\llbracket\psi\rrbracket_\gamma^g=1 = \Sigma\llbracket\psi\rrbracket_\alpha^g=1 \cup Q$ for some $Q \subseteq \Sigma\llbracket\psi\rrbracket_\alpha^g=1/2$. But given (a) and (b) it clearly follows that for all $Q \subseteq \Sigma\llbracket\psi\rrbracket_\gamma^g=1/2$ $\mathcal{P}(\Sigma\llbracket\psi\rrbracket_\gamma^g=1 \cup Q) \leq \llbracket x \rrbracket_\alpha^{g,\delta}$. And since $\llbracket x \rrbracket_\alpha^{g,\delta} = \llbracket x \rrbracket_\gamma^{g,\delta}$, we have our result.

(ii) Assume that $\llbracket Cr(\psi) \leq x \rrbracket_\alpha^{g,\delta} = 0$. We want to show $\llbracket Cr(\psi) \leq x \rrbracket_\gamma^{g,\delta} = 0$. Our assumption tell us that either $\llbracket x \rrbracket_\alpha^{g,\delta} \notin \mathbb{R}$, or $\llbracket x \rrbracket_\alpha^{g,\delta} \in \mathbb{R}$ and for every $Q \subseteq \Sigma\llbracket\psi\rrbracket_\alpha^g=1/2$, $\mathcal{P}(\Sigma\llbracket\psi\rrbracket_\alpha^g=1 \cup Q) \not\leq \llbracket x \rrbracket_\alpha^{g,\delta}$. In the former case the argument is trivial, since $\llbracket x \rrbracket_\alpha^{g,\delta} = \llbracket x \rrbracket_\gamma^{g,\delta}$. Considering the latter case, it once again follows from (a) and (b) that for every $Q \subseteq \Sigma\llbracket\psi\rrbracket_\gamma^g=1/2$ $\mathcal{P}(\Sigma\llbracket\psi\rrbracket_\gamma^g=1 \cup Q) \not\leq \llbracket x \rrbracket_\alpha^{g,\delta}$. And since $\llbracket x \rrbracket_\alpha^{g,\delta} = \llbracket x \rrbracket_\gamma^{g,\delta}$, we have our result.

Given, then, a classical model for \mathcal{L}^- , we can construct a VF model for \mathcal{L} via the Kripke construction. And given this ability to construct VF models for the language \mathcal{L} containing $Cr(\psi) \leq x$ and $Cr(\psi) \geq x$, we can construct VF^+ models for the language \mathcal{L}^+ in the manner described in §2.2.3. What we'll now do is use these models to show how a paracomplete theorist can resolve the paradox of credence.

Recall that we stipulated that Alpha's introspective abilities are such that following holds:

$$(9) \quad \forall x \in \mathbb{R} \quad (Cr(Cr(T(\neg\beta'\neg)) = x) = 1 \leftrightarrow Cr(T(\neg\beta'\neg)) = x)$$

We then showed that, given this assumption, Alpha is unable to satisfy the requirements imposed by:

$$\mathbf{CREDENCE}_1 \models \phi \rightarrow \psi \Rightarrow O(Cr(\phi) \leq Cr(\psi))$$

$$\mathbf{CREDENCE}_2 \quad O(Cr(\neg\phi) = 1 - Cr(\phi))$$

What we'll now show is that if we represent the facts about an agent's credences using a VF^+ model, such an agent can satisfy (9) while meeting the requirements imposed by $\mathbf{CREDENCE}_1$ and $\mathbf{CREDENCE}_2$.

To do this, however, we first need to do two things.

First, we need to add to our language the following symbols: $-$, 1 , \in , \mathbb{R} . $-$ will be interpreted in our language as the minus function. 1 will denote the number 1. \in will be interpreted as the set-membership relation. \mathbb{R} will be taken to denote the set of real numbers.

Second we need to provide *paraphrases* of $\mathbf{CREDENCE}_1$ and $\mathbf{CREDENCE}_2$. Both $\mathbf{CREDENCE}_1$ and $\mathbf{CREDENCE}_2$ as stated involve a binary relation symbol: $=$. But we don't have such a symbol in \mathcal{L}^+ . The occurrence of $=$ in $Cr(\phi) = x$ is really just a pneumatic device to indicate the appropriate interpretation. Despite not having sentences of the form: $Cr(\phi) = Cr(\psi)$ as a part of \mathcal{L}^+ , we can provide the following paraphrases:

$$\mathbf{CREDENCE}_1^* \models \phi \rightarrow \psi \Rightarrow O(\forall x(Cr(\phi) = x \rightarrow Cr(\psi) \geq x))$$

$$\mathbf{CREDENCE}_2^* \quad O(\forall x(Cr(\neg\phi) = x \leftrightarrow Cr(\phi) = 1 - x))$$

We can now prove that any agent whose credences are represented by a VF^+ model—call this a *VF⁺ agent*—will satisfy (9) together with all the requirements imposed by $\mathbf{CREDENCE}_1^*$ and $\mathbf{CREDENCE}_2^*$.

Proof that a VF^+ Agent Satisfies (9)

Let M be a VF^+ model for \mathcal{L}^+ . We'll show that for every point δ : $\llbracket \forall x \in \mathbb{R} (Cr(Cr(T(\Gamma\beta'\neg)) = x) = 1 \leftrightarrow Cr(T(\Gamma\beta'\neg)) = x) \rrbracket^\delta = 1$. To do this we show that in every Kripke model α in the construction of our VF^+ model: $\llbracket \forall x \in \mathbb{R} (Cr(Cr(T(\Gamma\beta'\neg)) = x) = 1 \leftrightarrow Cr(T(\beta')) = x) \rrbracket_\alpha^\delta = 1$. To show this we show that for every sequence g that assigns to x a member of \mathbb{R} , $\llbracket (Cr(Cr(T(\Gamma\beta'\neg)) = x) = 1) \rrbracket_\alpha^{g,\delta} = \llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta}$.

Here we argue by cases.

(i) Let $\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta} = 1$. It's easy enough to confirm by inspecting our clauses for specifying semantic values that, in VF^+ models, formulas of the form $Cr(\phi) = x$ have the same semantic value at every point of evaluation. Thus, for every $\delta' \in \Delta_\alpha$ $\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta'} = 1$. This tells us that $\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1} = 1} = \Delta_\alpha$. And so clearly for every $Q \subseteq \Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}}$ $\mathcal{P}(\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1}} \cup Q) = 1$. And so: $\llbracket (Cr(Cr(T(\Gamma\beta'\neg)) = x) = 1) \rrbracket_\alpha^{g,\delta} = 1$.

(ii) Let $\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta} = 0$. Again, it follows that for every $\delta' \in \Delta_m$ $\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta'} = 0$. Given this, we have: $\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1} = 1} = \Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}} = \emptyset$. And so for every $Q \subseteq \Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}}$, $\mathcal{P}(\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1}} \cup Q) = 0$. It follows that $\llbracket (Cr(Cr(T(\Gamma\beta'\neg)) = x) = 1) \rrbracket_\alpha^{g,\delta} = 0$.

(iii) Finally let $\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,\delta} = 1/2$. We want to show (a) that there is some $Q \subseteq \Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}}$ such that $\mathcal{P}(\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1}} \cup Q) = 1$ and (b) that there is some $Q' \subseteq \Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}}$ such that $\mathcal{P}(\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1}} \cup Q') \neq 1$. Since every formula of the form $Cr(\phi) = x$ has the same value at every point of evaluation, it follows that $\Sigma^{\llbracket Cr(T(\Gamma\beta'\neg)) = x \rrbracket_\alpha^{g,1/2}} = \Delta_\alpha$. But, then, (a) and (b) are obvious consequences of the fact that we can find a subset of Δ_α which has measure 1 and one which does not, e.g., Δ_α and \emptyset .

It follows, then, that in any VF^+ model (9) will be satisfied.

Proof that a VF^+ agent satisfies CREDENCE₁*

Assume that $\models \phi \rightarrow \psi$. We'll show that for every δ , $\llbracket \forall x (Cr(\phi) = x \rightarrow Cr(\psi) \geq x) \rrbracket^\delta = 1$. To show this we show that for every sequence g : $\llbracket Cr(\phi) = x \rightarrow Cr(\psi) \geq x \rrbracket^{g,\delta} = 1$. To show that this holds show that for every Kripke model α in the construction of our VF^+ model: $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} \leq \llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta}$.

We'll break this into two cases.

(i) Assume $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} = 1$. We'll show that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} = 1$. Our assumption tells us that for every $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha^{g,1/2}}$, $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^{g,1}} \cup Q) = \llbracket x \rrbracket_\alpha^g$. This, of course, ensures

that: $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1}) = \llbracket x \rrbracket_\alpha^g$. Given that $\models \phi \rightarrow \psi$ we know that $\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \subseteq \Sigma^{\llbracket \psi \rrbracket_\alpha^g=1}$. Thus for every $Q \subseteq \Sigma^{\llbracket \psi \rrbracket_\alpha^g=1/2}$, $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket_\alpha^g=1} \cup Q) \geq \llbracket x \rrbracket_\alpha^g$. This gives us that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} = 1$.

(ii) Assume $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} = 1/2$. We'll show that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} \geq 1/2$. Our assumption tells us that there are $Q, Q' \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha^g=1/2}$ such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q) = \llbracket x \rrbracket_\alpha^g$ and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q') \neq \llbracket x \rrbracket_\alpha^g$. Since $\models \phi \rightarrow \psi$ we know that $(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q) \subseteq (\Sigma^{\llbracket \psi \rrbracket_\alpha^g=1} \cup \Sigma^{\llbracket \psi \rrbracket_\alpha^g=1/2})$. Since $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q) = \llbracket x \rrbracket_\alpha^g$, it follows that there is some $Q'' \subseteq \Sigma^{\llbracket \psi \rrbracket_\alpha^g=1/2}$ such that $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket_\alpha^g=1} \cup Q'') \geq \llbracket x \rrbracket_\alpha^g$. This gives us that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} \geq 1/2$.

Our two cases show that in any VF^+ model we have $\llbracket \forall x(Cr(\phi) = x \rightarrow Cr(\psi) \geq x) \rrbracket_\alpha^\delta = 1$ for every δ , given that $\models \phi \rightarrow \psi$. This suffices to show that in any VF^+ model all of the requirements imposed by CRENCE_1^* will be satisfied.

Proof that a VF^+ agent satisfies CRENCE_2^*

We'll show that for every δ , $\llbracket \forall x(Cr(\neg\phi) = x \leftrightarrow Cr(\phi) = 1 - x) \rrbracket_\alpha^\delta = 1$. To show this we show that every Kripke model α in the construction of our VF^+ model is such that: $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = \llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta}$.

To see why this holds note the following facts:

(a) Δ_α can be partitioned into the following three sets: $\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1}$, $\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1}$ and $\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1/2}$. The latter set is identical to $\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1/2}$.

(b) For any $Q \subseteq (\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1} \cup \Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1/2})$, the complement of Q , Q^C is such that $Q^C \subseteq (\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup \Sigma^{\llbracket \phi \rrbracket_\alpha^g=1/2})$.

(c) For any $Q \subseteq (\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1} \cup \Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1/2})$, $\mathcal{P}(Q) = x$ iff $\mathcal{P}(Q^C) = 1 - x$.

Given (a)-(c) it is clear that:

(i) If every $Q \subseteq \Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1/2}$ is such that $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1} \cup Q) = x$, then every $Q' \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha^g=1/2}$ is such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q') = 1 - x$. Thus, if $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 1$, then $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 1$.

(ii) If every $Q \subseteq \Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1/2}$ is such that $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g=1} \cup Q) \neq x$, then every $Q' \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha^g=1/2}$ is such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g=1} \cup Q') \neq 1 - x$. Thus, if $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 0$, then $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 0$.

(iii) If there's a $Q, Q' \subseteq \Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g = 1/2}$ such that $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g = 1} \cup Q) = x$ and $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_\alpha^g = 1} \cup Q') \neq x$, then there's a $Q'', Q''' \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha^g = 1/2}$ such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g = 1} \cup Q'') = 1 - x$ and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha^g = 1} \cup Q''') \neq 1 - x$. Thus, if $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 1/2$, then $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 1/2$.

(i) - (iii) assure us that every Kripke model α in the construction of our VF^+ model is such that: $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = \llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta}$. And so we can be assured that: $\llbracket \forall x(Cr(\neg\phi) = x \leftrightarrow Cr(\phi) = 1 - x) \rrbracket^\delta = 1$. This suffices to show that in any VF^+ model all of the requirements imposed by CREDENCE_2^* will be satisfied.

In the class of models that paracomplete theorists will appeal to to characterize the extension of \models , $T(\Gamma\beta'\neg) \leftrightarrow \neg Cr(T(\Gamma\beta'\neg)) > 0.9$ will have semantic value 1 at every point of evaluation. What the above proofs show, then, is that if we are paracomplete theorists we can accept that $\models T(\Gamma\beta'\neg) \leftrightarrow \neg Cr(T(\Gamma\beta'\neg)) > 0.9$ and yet nonetheless resolve the paradox developed in §3.2 without rejecting either the set-up of the case or either of our plausible principles of credal rationality. In order, however, to do this, we need to allow that an agent's credential state may give rise to certain failures of classical logic. In particular, in any of the VF^+ models in which (9), CREDENCE_1^* and CREDENCE_2^* are all satisfied, excluded-middle will not hold for $\neg Cr(T(\Gamma\beta'\neg)) > 0.9$.

4.2.2 The Supervaluationist Solution

We've seen that there are VF^+ models in which Alpha's introspective powers are as characterized by (9), and in which Alpha is also able to meet all of the requirements imposed by CREDENCE_1^* and CREDENCE_2^* . We can also show that there are SV^+ models that have this feature as well. If, then, Alpha's credential state can give rise to the sorts of failures of classical logic that a supervaluationist thinks arise in cases of semantic paradox, Alpha need not be forced violate the rational requirements imposed by CREDENCE_1 and CREDENCE_2 .

Again, we'll let \mathcal{L} be a first-order language containing the following expressions: $Cr(\cdot) \leq x$ and $Cr(\cdot) \geq x$ (in addition to a truth-predicate, the normal first-order quantifiers and boolean connectives). An SV model for \mathcal{L} is (again) a tuple $\langle D_m, F, \Delta_m, \mathcal{P}, \llbracket \cdot \rrbracket_m \rangle$, where \mathcal{P} is a probability measure on Δ_m . Semantic values for formulas in an SV model are determined in the manner specified in §2.2.1, where formulas of the form: $Cr(\cdot) \leq x$ and $Cr(\cdot) \geq x$ are treated as complex. To determine how the value of formulas of the form: $Cr(\cdot) \leq x$ and $Cr(\cdot) \geq x$ are determined in such a model we simply need to specify how semantic values are assigned to such formulas in a classical model M .

Let M be a classical model $\langle D_m, F, \Delta_m, \mathcal{P}, \llbracket \cdot \rrbracket_m \rangle$. We say:

- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = Cr(\psi) \leq x$ then:
 $\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1}) \leq \llbracket x \rrbracket_m^{g,\delta}$.
 Otherwise $\llbracket \phi \rrbracket_m^{g,\delta} = 0$.
- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = Cr(\psi) \geq x$ then:

$\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1}) \geq \llbracket x \rrbracket_m^{g,\delta}$.

Otherwise $\llbracket \phi \rrbracket_m^{g,\delta} = 0$.

It is a consequence of these clauses, together with our definition of $Cr(\psi) = x$, that we have:

- For every wff ϕ and $g \in F$, and $\delta \in \Delta_m$, if $\phi = (Cr(\psi) = x)$ then:

$\llbracket \phi \rrbracket_m^{g,\delta} = 1$ iff $\llbracket x \rrbracket_m^{g,\delta} \in \mathbb{R}$ and $\mathcal{P}(\Sigma^{\llbracket \psi \rrbracket^g = 1}) = \llbracket x \rrbracket_m^{g,\delta}$.

Otherwise $\llbracket \phi \rrbracket_m^{g,\delta} = 0$.

The SV models that we'll be interested in will be those arrived at through the Kripke construction. (MONOTONICITY will still hold given the addition of $Cr(\cdot) \leq x$ and $Cr(\cdot) \geq x$ to our language, so such models will exist.) SV^+ models for \mathcal{L}^+ will be constructed in the manner outlined in §2.

What we want to do now is show that there are SV^+ models in which Alpha can satisfy (9) together with the requirements imposed by $CREDENCE_1^*$ and $CREDENCE_2^*$.

Proof that an SV^+ Agent Satisfies (9)

Let M be an SV^+ model for \mathcal{L}^+ . We'll show that for every point δ : $\llbracket \forall x \in \mathbb{R} (Cr(Cr(T(\beta')) = x) = 1 \leftrightarrow Cr(T(\beta')) = x) \rrbracket^\delta = 1$. To do this we show that in every Kripke model α in the construction of our SV^+ model: $\llbracket \forall x \in \mathbb{R} (Cr(Cr(T(\beta')) = x) = 1 \leftrightarrow Cr(T(\beta')) = x) \rrbracket_\alpha^\delta = 1$. To show this we show that for every sequence g that assigns to x a member of \mathbb{R} , $\llbracket (Cr(Cr(T(\beta')) = x) = 1) \rrbracket_\alpha^{g,\delta} = \llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta}$.

Here we argue by cases.

(i) Let $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta} = 1$. It's easy enough to confirm by inspecting our clauses for specifying semantic values that in SV^+ models formulas of the form $Cr(\phi) = x$ have the same semantic value at every point of evaluation. Thus, for every $\delta' \in \Delta_\alpha$ $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta'} = 1$. This tells us that for every classical closure α^\uparrow , $\Sigma^{\llbracket Cr(T(\beta')) = x \rrbracket_{\alpha^\uparrow}^g = 1} = \Delta_\alpha$. This means that for every classical closure α^\uparrow , $\llbracket (Cr(Cr(T(\beta')) = x) = 1) \rrbracket_{\alpha^\uparrow}^{g,\delta} = 1$. And so: $\llbracket (Cr(Cr(T(\beta')) = x) = 1) \rrbracket_\alpha^{g,\delta} = 1$.

(ii) Let $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta} = 0$. Again, it follows that for every $\delta' \in \Delta_m$ $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta'} = 0$. This means that for every classical closure α^\uparrow , $\Sigma^{\llbracket Cr(T(\beta')) = x \rrbracket_{\alpha^\uparrow}^g = 1} = \emptyset$. And so for every classical closure α^\uparrow , $\llbracket (Cr(Cr(T(\beta')) = x) = 1) \rrbracket_{\alpha^\uparrow}^{g,\delta} = 0$. It follows that $\llbracket (Cr(Cr(T(\beta')) = x) = 1) \rrbracket_\alpha^{g,\delta} = 0$.

(iii) Finally let $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta} = 1/2$. Again it follows that for every $\delta' \in \Delta_m$, $\llbracket Cr(T(\beta')) = x \rrbracket_\alpha^{g,\delta'} = 1/2$. This tells us:

(a) There is a classical closure α^\uparrow such that $\mathcal{P}(\Sigma^{\llbracket T(\beta') \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta} = \llbracket x \rrbracket_\alpha^{g,\delta}$.

(b) There is a classical closure α^\uparrow such that $\mathcal{P}(\Sigma^{\llbracket T(\beta') \rrbracket_{\alpha^\uparrow}^g = 1}) \neq \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta} = \llbracket x \rrbracket_\alpha^{g,\delta}$.

We want to show that $\llbracket (Cr(Cr(T(\beta') = x) = 1) \rrbracket_\alpha^{g,\delta} = 1/2$. This means that we want to show:

(a') There is a classical closure α^\uparrow such that $\mathcal{P}(\Sigma^{\llbracket Cr(T(\beta') = x) \rrbracket_{\alpha^\uparrow}^g = 1}) = 1$.

(b') There is a classical closure α^\uparrow such that $\mathcal{P}(\Sigma^{\llbracket Cr(T(\beta') = x) \rrbracket_{\alpha^\uparrow}^g = 1}) \neq 1$.

However, it is easy to see that (a) entails (a'), while (b) entails (b'). It follows, then, that in any SV^+ model (9) will be satisfied.

Proof that a SV^+ agent satisfies CREDCENCE_2^*

We'll show that for every δ , $\llbracket \forall x (Cr(\neg\phi) = x \leftrightarrow Cr(\phi) = 1 - x) \rrbracket^\delta = 1$. To show this we show that every Kripke model α in the construction of our SV^+ model is such that: $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = \llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta}$.

We know that for every classical closure α^\uparrow , $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha^\uparrow}^g = 1}) = x$ iff $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = 1 - x$. Using this fact, we argue by cases:

(i) Let $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 1$. We know that for every classical closure α^\uparrow , $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$ and so for every α^\uparrow , $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = 1 - \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. Thus $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 1$.

(ii) Let $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 0$. Thus for every classical closure α^\uparrow , $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha^\uparrow}^g = 1}) \neq \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. And so for every classical closure α^\uparrow , $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) \neq 1 - \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. And so $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 0$.

(iii) Let $\llbracket Cr(\neg\phi) = x \rrbracket_\alpha^{g,\delta} = 1/2$. Then there is some α^\uparrow such that, $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$, and some α^\uparrow such that, $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha^\uparrow}^g = 1}) \neq \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. But then clearly there is some α^\uparrow such that, $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = 1 - \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$, and some α^\uparrow such that, $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) \neq 1 - \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. Thus $\llbracket Cr(\phi) = 1 - x \rrbracket_\alpha^{g,\delta} = 1/2$.

This suffices to show that in any SV^+ model all of the requirements imposed by CREDCENCE_2^* will be met.

Proof that a SV^+ Agent Can Satisfy CREDCENCE_1^*

We'll now show that there are at least some SV^+ models in which an agent can satisfy all the demands imposed by CREDENCE_1^* . Since we have already seen that in every SV^+ model an agent will satisfy (9) and meet all the requirements imposed by CREDENCE_2^* , it follows that there are some SV^+ models in which an agent can satisfy (9) while meeting all of the demands imposed by CREDENCE_1^* and CREDENCE_2^* .

Assume that $\models \phi \rightarrow \psi$. We'll show that there are SV^+ models such that for every δ , $\llbracket \forall x(Cr(\phi) = x \rightarrow Cr(\psi) \geq x) \rrbracket^\delta = 1$. To show this we show that there are models such that for every sequence g : $\llbracket Cr(\phi) = x \rightarrow Cr(\psi) \geq x \rrbracket^{g,\delta} = 1$. And to show that this holds we show that there are SV^+ models such that for every Kripke model α in the construction of the SV^+ model (after the point at which the construction has stabilized): $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} \leq \llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta}$.

We'll break this into two cases.

(i) In this first type of case, $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} = 1$. This means that for every classical closure α^\uparrow $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}$. We want to show that it follows from our assumption that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} = 1$. To do this it will suffice to show:

$$(A) \models \phi \rightarrow \psi \Rightarrow [(\forall \alpha^\uparrow \mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}) \Rightarrow (\forall \alpha^\uparrow \mathcal{P}(\Sigma^{\llbracket \psi \rrbracket_{\alpha^\uparrow}^g = 1}) \geq \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta})].$$

To show that this holds it would suffice to show:

$$(B) \models \phi \rightarrow \psi \Rightarrow [\forall \gamma^\uparrow, Q[(Q = \{\delta : \llbracket \psi \rrbracket_{\gamma^\uparrow}^\delta = 1\}) \Rightarrow \exists \alpha^\uparrow, Q'(Q' = \{\delta : \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 1\} \wedge Q' \subset Q)].$$

And to show that the latter claim holds it would suffice to show that:

$$(C) \forall Q \subseteq \Delta_\alpha [(\forall \delta \in Q \exists \alpha^\uparrow \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 0) \Rightarrow (\exists \alpha^\uparrow \forall \delta \in Q \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 0)].$$

To see why (C) entails (B), let $\models \phi \rightarrow \psi$ and assume that $Q = \{\delta : \llbracket \psi \rrbracket_{\gamma^\uparrow}^\delta = 1\}$ for some γ^\uparrow . $Q^C = \{\delta : \llbracket \psi \rrbracket_{\gamma^\uparrow}^\delta = 0\}$. Since $\phi \rightarrow \psi$ we know that $\forall \delta \in Q^C \exists \alpha^\uparrow \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 0$. By (C) we have $\exists \alpha^\uparrow \forall \delta \in Q^C \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 0$. But this guarantees that $\exists \alpha^\uparrow, Q'(Q' = \{\delta : \llbracket \phi \rrbracket_{\alpha^\uparrow}^\delta = 1\} \wedge Q' \subset Q)$.

(ii) In the second type of case, $\llbracket Cr(\phi) = x \rrbracket_\alpha^{g,\delta} = 1/2$. We want to show that it follows from our assumption that $\llbracket Cr(\psi) \geq x \rrbracket_\alpha^{g,\delta} \geq 1/2$. To do this it will suffice to show:

$$(A') \models \phi \rightarrow \psi \Rightarrow [(\exists \alpha^\uparrow \mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha^\uparrow}^g = 1}) = \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta}) \Rightarrow (\exists \alpha^\uparrow \mathcal{P}(\Sigma^{\llbracket \psi \rrbracket_{\alpha^\uparrow}^g = 1}) \geq \llbracket x \rrbracket_{\alpha^\uparrow}^{g,\delta})].$$

To show that this holds it would suffice to show:

$$(B') \models \phi \rightarrow \psi \Rightarrow [\forall \gamma^\uparrow, Q[(Q = \{\delta : \llbracket \phi \rrbracket_{\gamma^\uparrow}^\delta = 1\}) \Rightarrow \exists \alpha^\uparrow, Q'(Q' = \{\delta : \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1\}) \wedge Q \subset Q']]].$$

And to show that the latter claim holds it would suffice to show that:

$$(C') \forall Q \subseteq \Delta_\alpha [(\forall \delta \in Q \exists \alpha^\uparrow \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1) \Rightarrow (\exists \alpha^\uparrow \forall \delta \in Q \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1)].$$

To see why (C') entails (B'), let $\models \phi \rightarrow \psi$ and assume that $Q = \{\delta : \llbracket \phi \rrbracket_{\gamma^\uparrow}^\delta = 1\}$ for some γ^\uparrow . Since $\models \phi \rightarrow \psi$ we know that $\forall \delta \in Q \exists \alpha^\uparrow \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1$. By (C') we can infer that $\exists \alpha^\uparrow \forall \delta \in Q \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1$. And from this it clearly follows that $\exists \alpha^\uparrow, Q'(Q' = \{\delta : \llbracket \psi \rrbracket_{\alpha^\uparrow}^\delta = 1\}) \wedge Q \subset Q'$.

What (i) and (ii) tell us is that we can be assured that there is an SV^+ model in which Alpha can satisfy the requirements imposed by CREDENCE_1^* , if we can be assured that there are SV^+ models such that (C) and (C') hold for every Kripke model α in its construction. But it's obvious that there are such models. Perhaps the clearest case is presented by any SV^+ model in which there is only a single member of Δ . But there will also be plenty of other more complex models meeting this condition. Given this we can be assured that there are models in which the strictures imposed by CREDENCE_1^* are met.

The supervaluationist solution is in certain respects clearly weaker than the paracomplete solution. In the paracomplete case we were able to show that in any VF^+ model for \mathcal{L}^+ , (9) would hold and all of the requirements of CREDENCE_1^* and CREDENCE_2^* would be met. In the supervaluationist case, however, all that we've shown is that this holds in a certain subset of SV^+ models.

Now this is enough to provide a solution to the paradox as it is presented. If we are supervaluationists and we allow that Alpha's credal state may give rise to certain failures of classical logic, then we can show that there are ways that Alpha's credal state could be, compatible with our stipulations about Alpha, in which Alpha is rationally blameless.

However, the fact that we haven't shown that in every such model (9), CREDENCE_1^* and CREDENCE_2^* will all hold leaves open the worry that we may be able to add to our description of Alpha in a way that (a) leaves Alpha antecedently rational but (b) precludes our characterizing Alpha's credal state in such a way that all the requirements imposed by CREDENCE_1^* and CREDENCE_2^* hold. So far as I can tell right now this is an open worry. Perhaps a more general proof can be offered to show that in *every* SV^+ model (9), CREDENCE_1^* and CREDENCE_2^* will all hold. I don't know of any counterexample to this claim, but I also can't see a way to prove it. I'll leave this as an open question. It may be that a supervaluationist doesn't have the same resources as a paracomplete theorist for resolving the paradox of credence.

4.2.3 The Paraconsistent Solution

We've already seen that a VF^+ agent can in principle meet all of the requirements imposed by CREDENCE_1 and CREDENCE_2 , despite having the introspective powers attributed in (9). This tells us that both paracomplete and paraconsistent theorists can resolve the paradox of credence.

Of course, what this means for a paraconsistent theorist is very different from what it means for a paracomplete theorist. In the VF^+ models in which Alpha will satisfy (9) and, in addition, meet all of the requirements imposed by $CREDENCE_1$ and $CREDENCE_2$, $Cr(T(\ulcorner\beta'\urcorner)) \geq 0.9$ will have semantic value $1/2$ and so will $\neg Cr(T(\ulcorner\beta'\urcorner)) \geq 0.9$. Since the paraconsistent theorist takes $1/2$ to be a designated value, it follows that if Alpha's credal state is to satisfy the rational requirements imposed by $CREDENCE_1$ and $CREDENCE_2$ in this manner, then Alpha's credence must both be greater than 0.9 and not greater than 0.9 . The paraconsistent theorist can hold, then, that Alpha can indeed meet all the requirements imposed by $CREDENCE_1$ and $CREDENCE_2$. However, this requires allowing that the same type of contradictions that arise in cases of semantic paradox arise from an agent's credal state.

Of course, as in the case of qualitative belief, it might seem that a paraconsistent theorist could offer a simpler solution to this paradox. Shouldn't the paraconsistent theorist just reject $CREDENCE_2$? After all, the paraconsistent theorist will presumably think that there are claims ϕ such that an agent should have credence greater than 0.5 in ϕ and in $\neg\phi$. For example, the paraconsistent theorist will presumably have pretty high credence in both the proposition expressed by the liar sentence $\ulcorner\lambda\urcorner$ and in the proposition expressed by its negation $\ulcorner\neg\lambda\urcorner$. Let's say that the paraconsistent theorist has 0.9 credence in both the proposition expressed by $\ulcorner\lambda\urcorner$ and in the proposition expressed by $\ulcorner\neg\lambda\urcorner$. Given that these are her credences, shouldn't the paraconsistent theorist deny that there is a rational obligation to have, say, 0.1 credence in the negation of a proposition in which one has credence 0.9 ?

Here, again, I think that there are some subtle issues, similar to the issues that arise over whether the paraconsistent theorist should reject $CONSISTENCY$. As in the earlier case, the paraconsistent theorist does not *need* to reject $CREDENCE_2$ in order to resolve the paradox. And, I'll argue later that, in fact, there are good reasons for the paraconsistent theorist to endorse $CREDENCE_2$. However, as in the earlier case, this issue does not really matter for our present purposes.

The main point that I want to make in this section is that a paraconsistent theorist can resolve the paradox of credence, but doing so requires allowing that in certain cases a rational agent's credal state must give rise to the same failures of classical logic as arise in cases of semantic paradox.

We've already seen how Alpha can meet the requirements imposed by $CREDENCE_1$ and $CREDENCE_2$, if Alpha's credence in $T(\ulcorner\beta'\urcorner)$ is both greater than and not greater than 0.9 . And it's easy to see that there is no way, by paraconsistent lights, for Alpha to meet these requirements without there being such credal dialethias. For the only way, by paraconsistent lights, to block the argument given in §3.2 for the conclusion that Alpha cannot meet the requirements imposed by $CREDENCE_1$ and $CREDENCE_2$, is if one allows that an agent's having certain credences does not preclude her *not* having those credences.

So, if the paraconsistent theorist wants to hold on to $CREDENCE_1$, $CREDENCE_2$ and $POSSIBILITY$, then she should allow that an agent's credal state may be a source of dialethias. In particular, the paraconsistent theorist should allow that the claim $Cr(T(\ulcorner\beta'\urcorner)) > 0.9$ will be a dialethia.

If, on the other hand, the paraconsistent theorist wishes to resolve the paradox by rejecting the relevant instances of $CREDENCE_2$, we can still argue that she must allow that in such cases an

agent's credal state may be a source of dialethias. The argument for this claim is really the same as the argument for the parallel claim in the case of qualitative beliefs. While a paraconsistent theorist may not want to accept CREDENCE_2 as an unrestricted principle covering all propositions, she will presumably want to accept it in those cases in which there are no dialethias. What this means is that if the paraconsistent theorist is to block the paradox by rejecting the particular instances of CREDENCE_2 to which appeal is made in §3.2 then she will need to allow that $Cr(T(\ulcorner\beta'\urcorner)) > 0.9$ is a dialethia.

In either case, then, we get the conclusion that resolving the paradox requires allowing that a rational agent must have a credal state that gives rise to true contradictions.

4.3 Non-classical Solutions to the Paradox of Knowledge

4.3.1 The Paracomplete Solution

In §3.3 we saw that, given classical logical reasoning, in certain cases an agent could not come to know a proposition that was deduced by a known valid method.

The paracomplete theorist certainly has resources to block the argument to this conclusion. For this argument relies on the claim that *reductio* reasoning is valid. That is, the argument relies on the validity of the following meta-rule: $\phi \models \neg\phi \Rightarrow \models \neg\phi$. While this is certainly valid in classical logic it is not valid in the logic endorsed by paracomplete theorists. To see this, note that a paracomplete theorist will endorse: $T(\ulcorner\lambda\urcorner) \models \neg T(\ulcorner\lambda\urcorner)$. But the paracomplete theorist will not endorse: $\models \neg T(\ulcorner\lambda\urcorner)$.

So the paracomplete theorist, then, can block the argument given in §3.3 for the incompatibility of **FACTIVITY** and **DEDUCTION**. What the paracomplete theorist can say is that although we have $KT(\ulcorner\kappa\urcorner) \models \neg KT(\ulcorner\kappa\urcorner)$, this doesn't entail $\neg KT(\ulcorner\kappa\urcorner)$. In this way, the paracomplete theorist can avoid the conclusion that an agent may knowingly deduce conclusions that the agent, nonetheless, cannot come to know. Importantly, however, for this to work the paracomplete theorist must allow that excluded-middle may fail for the claim that Alpha knows that $\ulcorner\kappa\urcorner$ is true. For as long as excluded-middle holds for the relevant claim, $\neg KT(\ulcorner\kappa\urcorner)$ will follow from $KT(\ulcorner\kappa\urcorner) \models \neg KT(\ulcorner\kappa\urcorner)$.

The paracomplete theorist, then, can resolve the paradox of knowledge, but, once again, doing so requires allowing that an agent's propositional attitudes, in this case an agent's knowledge state, may give rise to the same failures of classical logic as the paracomplete theorist thinks arise in cases of semantic paradox.

And, as in the case of the paradox of belief, the paracomplete theorist can do more than merely reject a certain step in the development of our paradox. The paracomplete theorist can provide a class of models that guarantee that **FACTIVITY** and **DEDUCTION** hold.

Let \mathcal{L} be a first-order language with an operator K . We let \mathcal{M} be the class VF^+ models for \mathcal{L} in which R_m is reflexive, and K is treated in the models like \Box in the models developed in §2.2. We take the validity relation for \mathcal{L} to be defined as preservation of value 1 in this class of models. We can show that the following all hold:

FACTIVITY $K\phi \models \phi$

DEDUCTION' $\models \phi \Rightarrow \models K\phi$

KAPPA $\models T(\ulcorner \kappa \urcorner) \leftrightarrow \neg KT(\ulcorner \kappa \urcorner)$

What this means is that if we are paracomplete theorists we can allow that knowledge is factive and that an agent may know every proposition for which there is a valid derivation, despite the validity of $T(\ulcorner \kappa \urcorner) \leftrightarrow \neg KT(\ulcorner \kappa \urcorner)$. Given that DEDUCTION' holds, DEDUCTION will hold *a fortiori*.

Proof of FACTIVITY

We let α be a Kripke model at some stage in the construction of our VF^+ model M . We'll show that in any such model for every point $\delta \in \Delta$ if $\llbracket K\phi \rrbracket_\alpha^\delta = 1$ then $\llbracket \phi \rrbracket_\alpha^\delta = 1$. The proof of this is trivial. Let $\llbracket K\phi \rrbracket_\alpha^\delta = 1$. By our semantic clauses this means that for every δ' such that $\delta R\delta'$ $\llbracket \phi \rrbracket_\alpha^{\delta'} = 1$. And since R is reflexive, we have $\delta R\delta$. And so $\llbracket \phi \rrbracket_\alpha^\delta = 1$.

Since we've established that for every Kripke model in the construction of VF^+ model M is such that if $\llbracket K\phi \rrbracket_\alpha^\delta = 1$ then $\llbracket \phi \rrbracket_\alpha^\delta = 1$, it follows that if $\llbracket K\phi \rrbracket_m^\delta = 1$ then $\llbracket \phi \rrbracket_m^\delta = 1$.

Proof of DEDUCTION'

The proof of DEDUCTION' is even more trivial. Assume that $\models \phi$. Then we have that for any VF^+ model M in our set, and any point δ , $\llbracket \phi \rrbracket_m^\delta = 1$. But then clearly in any model M , for any point δ , for every δ' such that $\delta R\delta'$, we have $\llbracket \phi \rrbracket_m^{\delta'} = 1$. Thus, given that we have $\models \phi$ it follows that in any model M for any point δ we have $\llbracket K\phi \rrbracket_m^\delta = 1$, and so $\models K\phi$.

Proof of KAPPA

We assume that in the class of VF^+ models under consideration $\llbracket \ulcorner \kappa \urcorner \rrbracket = \neg KT(\ulcorner \kappa \urcorner)$. Given this fact it is easy to verify that in every Kripke model α in the construction of our VF^+ models, at every point δ , $\llbracket T(\ulcorner \kappa \urcorner) \rrbracket_\alpha^\delta = \llbracket \neg KT(\ulcorner \kappa \urcorner) \rrbracket_\alpha^\delta = 1/2$. So, for every Kripke model α (after the first stage) in the construction of our VF^+ models, at every point δ , $\llbracket T(\ulcorner \kappa \urcorner) \leftrightarrow \neg KT(\ulcorner \kappa \urcorner) \rrbracket_\alpha^\delta = 1$. And so in every VF^+ model under consideration at every point δ , $\llbracket T(\ulcorner \kappa \urcorner) \leftrightarrow \neg KT(\ulcorner \kappa \urcorner) \rrbracket^\delta = 1$.

We can be assured, then, that knowledge can be factive and that, in principle, an agent such as Alpha can come to know whatever it has knowingly deduced, despite the existence of a sentence such as $\ulcorner \kappa \urcorner$. However, in the models in which these facts are compatible, excluded-middle will fail for the claim that Alpha knows that $\ulcorner \kappa \urcorner$ is true. Indeed, in these models it will be indeterminate whether Alpha knows this. The paracomplete theorist can resolve our epistemic paradox, but doing so requires that facts about whether or not Alpha knows certain propositions may be subject to failures of classical logic.

4.3.2 The Supervaluationist Solution

A supervaluationist will also reject the meta-rule that allows one to infer $\models \neg\phi$ from $\phi \models \neg\phi$. For this reason the supervaluationist can also block the paradox of knowledge developed in §3.3.

The supervaluationist can also provide a class of SV^+ models in which FACTIVITY, DEDUCTION' and KAPPA all hold. Indeed, the proofs are trivial variants of those offered for VF^+ models. These are left to the reader. As in the paracomplete case, in the supervaluationist case it will turn out that in each model in which these conditions all hold, $\ulcorner\kappa\urcorner$ will have semantic value $1/2$. The supervaluationist can resolve the paradox of knowledge, but doing so requires that an agent's epistemic state give rise to the same types of failure of classical logic as arise in the case of semantic paradox.

4.3.3 The Paraconsistent Solution

The paradox of knowledge is easily resolved by a paraconsistent theorist. The paraconsistent theorist can agree that Alpha doesn't know that $\ulcorner\kappa\urcorner$ is true. But this, of course, need not preclude Alpha from also knowing that $\ulcorner\kappa\urcorner$ is true. And so the derivation need not tell us that FACTIVITY and DEDUCTION don't hold.

Since the paraconsistent theorist employs the same model theory as the paracomplete theorist, our proof that there are classes of VF^+ models in which FACTIVITY, DEDUCTION' and KAPPA all hold, show us that this type of solution can be worked out in detail. In any such model, we noted that the claim that Alpha knows that $\ulcorner\kappa\urcorner$ is true will have semantic value $1/2$. By paraconsistent lights, what this means is that in order to allow for FACTIVITY and DEDUCTION to hold, we need to allow that Alpha both knows and doesn't know that $\ulcorner\kappa\urcorner$ is true.

Once again our rational paradoxes can be resolved if we allow for similar failures of classical logic as arise in cases of semantic paradoxes to arise also for claims about the epistemic states of certain agents.

Chapter 5

Non-Classical Attitudes

We've seen that *if* we allow that Alpha's doxastic/credal/epistemic states may be sources of certain failures of classical logic, then we can resolve the paradoxes developed in chapter 3. Of course, this would do nothing to resolve our paradoxes if doxastic/credal/epistemic states couldn't be sources of such failures of classical logic. *One* reason to hold that this is the case is a general rejection of the claim that classical logic can fail in any domain. This is not an unreasonable position, but it isn't one that we're going to be interested in here. Instead, what we want to know is whether or not one who accepts a paracomplete/paraconsistent/supervaluationist account of the semantic paradoxes should allow that such failures of classical logic can extend to the doxastic/credal/epistemic case.

In what follows, we'll first discuss the doxastic and credal cases, since these raise issues not raised by the epistemic case. It will turn out that there are some views about the nature of doxastic/credal states that potentially create trouble for the hypothesis that Alpha's doxastic and credal states can give rise to the appropriate failures of classical logic. We'll show, however, that there are other attractive views about the nature of these mental states that make such failures of classical logic seem quite natural given what has been established so far.

We'll then discuss the case for non-classicality in an agent's epistemic state.

5.1 Non-Classical Doxastic and Credal States

Should one allow that certain principles of classical logic may fail for claims about what an agent believes (or what credence an agent has in a proposition), if one allows for such failures in the case of semantic paradox? How one answers this question may depend in part on what commitments one has about the nature of doxastic/credal states.

I'm going to assume that physicalism is true, at least with respect to doxastic/credal states. In particular, I'm going to assume that every doxastic/credal state is identical to some physical or functional state of the agent.¹ There are different versions of physicalism. One could, for example,

¹To accommodate externalist intuitions about content, these claims may need to be qualified. For example, if one is convinced by certain externalist thought-experiments, then one may want to hold that belief states are to be reduced

hold that doxastic/credal states are **type** identical to some physical or functional state. A type identity theory says that given a doxastic state-type D there exists some, say, functional type F such that every instance of D is identical to an instance of F and vice versa. (This type of theory can come in various strengths, depending on how strong a modal connection the theory takes there to be between such state-types. At one extreme the theory could hold that it is metaphysically necessary that every instance of D is an instance of F and every instance of F is an instance of D . At the other extreme the theory could simply be a claim about the way things are in our world. And of course there are intermediate claims, e.g., that such a connection is nomological.) Or one could hold that such states are **token** identical to some physical or functional state. A token identity theory holds that given a doxastic state-type D , for every instance of D , there is some, say, functional type F such that the instance of D is identical to an instance of F . (Again, this type of theory can come in varying degrees of strength depending on what sort of modal status the proponent takes this claim to have.) For our purposes, it won't matter which is true. All that we'll require for what follows is the claim, endorsed by both theories, that every particular belief state is identical to some physical or functional state.²

One way of categorizing accounts of doxastic/credal states is in terms of what types of psychologically realized bearers of content the account requires. On a **content atomist** view, given a true belief ascription, e.g., *Sally believes that snow is white*, there will be some psychologically realized vehicle that bears just the content ascribed, viz., that snow is white.³ Similarly for ascriptions of credence. On a **content holist** view, such ascriptions may be true even though there is no psychologically realized vehicle with just that content.⁴ On this view, a doxastic state simply carves out the space of possible worlds, while a credal state simply determines a probability measure over the space of possible worlds. In virtue of an agent's having such a doxastic state, certain ascriptions of belief will be true. They will be true just in case the proposition that one ascribes is true in all of the worlds that are compatible with the agent's doxastic state. In virtue of the agent having such a credal state, certain statements about the agent's credence in a proposition will be true. They will be true in virtue of the measure on the set of worlds in which the proposition is true.

What I'll now argue is that content atomist views *may* preclude non-classical credal and doxas-

to some functional or neural state of the agent, F , plus some external facts about the agent's environment, E . This qualification won't matter for what follows. In the arguments below simply replace occurrences of ' F ' by ' F and E '.

²It is worth noting that the type of physicalism assumed here isn't really essential for the argument, although it provides a useful (and I think plausible) constraint. One way, already noted, in which this assumption could be dropped without harm to the argument is if we allowed external environmental conditions to play a role in determining an agent's doxastic state. But what is really essential to the argument is that the doxastic states of an agent are reducible to some domain such that it is independently implausible that certain claims about the reducing domain give rise to failures of classical logic. *One* way in which this could hold is if our version of physicalism is true, but there are others.

³See e.g., Fodor (1975).

⁴See Lewis (2000), Lewis (1999), Stalnaker (1984). A note of warning about the terminology. Sometimes when people talk about an "atomistic" view of content, what is meant is the view that it is possible for an agent have a belief with a certain content S and no other beliefs, while by a "holistic view of content", what is meant is the view that an agent, if she has some belief with content S , must also have other beliefs, with other contents. As should be clear, this is *not* how I am using these terms.

tic states (regardless of whether one endorses a paracomplete, paraconsistent, or supervaluationist logic.) On the other hand, content holist views impose no serious obstacle to doxastic and credal states giving rise to such failures of classical logic. Indeed, such failures of classical logic would seem to follow naturally from the combination of content holism together with the endorsement of a paracomplete/paraconsistent/supervaluationist logic.

5.1.1 Atomistic Contents and Non-Classical Doxastic/Credal States

I'll begin by outlining a general argument that can be used to create trouble for the combination of a content atomist view and non-classical doxastic/credal states, regardless of whether one endorses KFS^+ , LP^+ or SV^+ . The argument will focus on doxastic states, but it should be obvious how the same arguments could be run *mutatis mutandis* for credal states. I'll then consider how this type of argument might be resisted.

Finally, I'll consider an argument that creates trouble, in particular, for the existence doxastic/credal states that give rise to the failures of classical logic posited by paracomplete theories.

The First Argument

Our first argument goes as follows:

- (P1) $B_\alpha T(\ulcorner\beta\urcorner)$ cannot have any of the non-classical statuses that are represented (in either SV^+ or VF^+ models) by values other than $1/2$.
- (P2) If the theory of atomistic psychological contents is correct, $B_\alpha T(\ulcorner\beta\urcorner)$ cannot have the non-classical status that is represented (in either SV^+ or VF^+ models) by semantic value $1/2$.
- (C) If the theory of atomistic psychological contents is correct, $B_\alpha T(\ulcorner\beta\urcorner)$ cannot have any non-classical status.

Let me first provide a defense of (P1). I'll then provide arguments that seek to establish (P2) for paracomplete, paraconsistent and supervaluationist theories.

Defense of (P1)

We've seen that in order for Alpha to meet the requirements imposed by CONSISTENCY and EVIDENCE, the claim that Alpha believes that $\ulcorner\beta\urcorner$ is true must have a non-classical status. As it turns out, there are *many* distinct statuses that $B_\alpha T(\ulcorner\beta\urcorner)$ could have that would be compatible with meeting this demand. Such statuses are represented in our VF^+ and SV^+ models by semantic values other than 1 and 0. Nonetheless, there is good reason to hold that insofar as $B_\alpha T(\ulcorner\beta\urcorner)$ can have one of these non-classical statuses, it must have the status represented by semantic value $1/2$, i.e., the value that a formula has when it has value $1/2$ at every Kripke model in the construction of the VF^+ or SV^+ model. Here's why.

In the classes of non-classical models that define our paracomplete, paraconsistent and super-valuationist theories, the only claims that take values other than 1, 1/2 or 0 are claims involving conditionals.⁵ Claims that can be formulated in the fragment of the language excluding \rightarrow will all have value 1, 1/2 or 0. The reason for this is that such formulas will have the same value in every Kripke model in the construction of the VF^+ or SV^+ model. Given this fact, it is reasonable to hold that the non-classical statuses that are represented by values other than 1/2 in our VF^+ or SV^+ models are statuses that only claims involving conditionals can have. And since $B_\alpha T(\ulcorner\beta\urcorner)$ does not involve a conditional, it cannot have any of these non-classical statuses.

Defense of (P2)

I'll now offer a defense of (P2). The argument will be broken into cases. I'll first show that for a paracomplete or paraconsistent theory, a content atomist view precludes $B_\alpha T(\ulcorner\beta\urcorner)$ from having semantic value 1/2. Next, I'll show how the same conclusion follows for a supervaluationist theory.

Our arguments here will rely on the following auxiliary premiss, discussion of which I defer until later:

(P2') Whether an agent is in a certain functional (or brain) state at a certain time is not something that can be the source of failures of classical logic.

Atomistic Content and Paracomplete/Paraconsistent Theories

Let αRf mean *Alpha is in functional state f* , where f is a variable.⁶ We'll assume that this formula is satisfied only if f is assigned a functional state.

Given the truth of physicalism, if we endorse a content atomist theory then we'll have the following:

$$\exists f(B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow \alpha Rf)^7$$

That is, if one is a content atomist, given the truth of physicalism, one will be committed to the claim that there is some functional state f such that Alpha believes that $\ulcorner\beta\urcorner$ is true just in case Alpha is in functional state f .

We assume:

$$(i) \llbracket B_\alpha T(\ulcorner\beta\urcorner) \rrbracket = 1/2$$

⁵Terminological note: As I am using the term a formula *involves* a conditional if it either contains a conditional or contains a term that refers to a formula that involves a conditional.

⁶Nothing here will turn on it being a functional state, as opposed to, say, a neural state, that we're taking to be identical to the particular belief state.

⁷To forestall potential confusion, I note that this claim, as with all of our earlier claims about Alpha's doxastic state, is implicitly time-indexed. We could make this explicit by saying: $\exists f(B_\alpha tT(\ulcorner\beta\urcorner) \leftrightarrow \alpha Rtf)$, where t here is some term referring to a particular time. To see why both type and token identity theories are committed to this claim, note that a type identity theory will endorse the following: $\exists f\forall t(B_\alpha tT(\ulcorner\beta\urcorner) \leftrightarrow \alpha Rtf)$, while a token identity theory will endorse the weaker: $\forall t\exists f(B_\alpha tT(\ulcorner\beta\urcorner) \leftrightarrow \alpha Rtf)$. Our particular claim follows from each of these more general claims.

(ii) For every assignment function g , either $\llbracket \alpha Rf \rrbracket^g = 1$ or $\llbracket \alpha Rf \rrbracket^g = 0$.

(i) is assumed for the purpose of *reductio*. (ii) is justified by (P2').

We can show that from (i) and (ii), it follows that in our VF^+ model M , $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket = 0$.

We want to show that for every g , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket^g = 0$. But given (i) and (ii), it is obvious that this will hold. By (i), we know that for every g , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket^g = 1/2$. By (ii) we know that for every g , $\llbracket \alpha Rf \rrbracket^g = 1$ or $\llbracket \alpha Rf \rrbracket^g = 0$. In either case, if $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket^g = 1/2$ and $\llbracket \alpha Rf \rrbracket^g = 1$, or if $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket^g = 1/2$ and $\llbracket \alpha Rf \rrbracket^g = 0$, it will follow that $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket^g = 0$. This guarantees, then, that from (i) and (ii) it follows that $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket = 0$.

Since a content atomist will be committed to the truth of $\exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf)$, it follows that either (i) or (ii) must fail. Assuming, then, that whether an agent is in a certain functional state cannot be a matter for which classical logic fails, it follows that, whether one is a paracomplete or paraconsistent theorist, if one endorses content atomism, then one cannot allow $B_\alpha T(\ulcorner \beta \urcorner)$ to have the non-classical status represented by semantic value $1/2$ in VF^+ models.

Atomistic Content and Supervaluationist Theories

We can show that the same conclusion follows for a supervaluationist theory.

Let M be an SV^+ model. We again assume:

(i) $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket = 1/2$

(ii) For every assignment function g , either $\llbracket \alpha Rf \rrbracket^g = 1$ or $\llbracket \alpha Rf \rrbracket^g = 0$.

We'll show that from (i) and (ii) it follows that in M , $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket = 0$. Since a content atomist will be committed to the claim: $\exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf)$, it follows that, given (ii), $B_\alpha T(\ulcorner \beta \urcorner)$ cannot have semantic value $1/2$.

What we want to show is that once our model M has settled into its cyclic pattern, for every stage γ , $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket_\gamma = 0$. To show this, we want to show that for every classical closure γ^\uparrow , $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket_{\gamma^\uparrow} = 0$. To show that this holds, it suffices to show that for every assignment function g , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_{\gamma^\uparrow}^g = 0$. To show that this holds, it suffices to show that for every assignment function g , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_\gamma^g = 0$.

It is easy to establish that this latter claim follows from (i) and (ii). We argue by cases. We show that this holds if γ is a successor ordinal, i.e., if there is some α such that $\gamma = \alpha + 1$. We then show that this holds if γ is a limit ordinal.

Assume that γ is a successor ordinal, i.e., $\gamma = \alpha + 1$. By (i) and (ii), we know that for every assignment function g , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket_\alpha^g \neq \llbracket \alpha Rf \rrbracket_\alpha^g$. It follows, then, that $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_{\alpha+1}^g = 0$, i.e., $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_\gamma^g = 0$.

Assume next that γ is a limit ordinal. Since we've shown that for all the non-limit ordinals α preceding γ , $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_\alpha^g = 0$, it follows that $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf \rrbracket_\gamma^g = 0$.

Given assumptions (i) and (ii), it follows that in the SV^+ model M , $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket = 0$. Since a content atomist will be committed to the truth of $\exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf)$, it follows that

either (i) or (ii) must fail. Assuming, then, that whether an agent is in a certain functional state cannot be a matter for which classical logic fails, it follows that, if one is a supervaluationist and one endorses content atomism, then one cannot allow $B_\alpha T(\ulcorner\beta\urcorner)$ to have the non-classical status represented by semantic value $1/2$ in SV^+ models.

What we've argued is that whether one endorses a paracomplete, paraconsistent, or supervaluationist account, if one endorses an atomistic theory of content, then one cannot allow $B_\alpha T(\ulcorner\beta\urcorner)$ to have semantic value $1/2$, i.e., (P2) holds.

In order to block this argument for the claim that an atomistic theory of content precludes the type of failures of classical logic that our resolution of doxastic and credal paradoxes requires, one must either reject (P1) or (P2). In the latter case, the only real place to challenge the argument is (P2'). So in order to block this argument one must either reject (P1) or (P2').

Let's consider these options in reverse order.

Rejecting (P2')

Viewed in the most neutral light possible, one can see the above argument as showing that we need to give up *one* of the following claims: (i) the proposition that Alpha believes that $\ulcorner\beta\urcorner$ is true can have a non-classical status associated with semantic value $1/2$, (ii) for any proposition ϕ that an agent believes, there is a psychologically realized state that has just that content, (iii) claims about whether an agent is in a certain functional or neural state are not something that can be the source of failures of classical logic. In principle, one could leverage the above argument against any of these three claims. In particular, then, one could take the argument to show that (iii), i.e., (P2'), is false.

However, this seems like a bizarre way to twist this argument. Initially it is, I think, pretty plausible that claims about functional states (or brain states) just aren't the type of things that can give rise to the appropriate failures of classical logic that either the paracomplete, paraconsistent or supervaluationist theories endorse. At the very least, this would seem more plausible initially than (i) or (ii). Since we should try to hold on to those claims that we take to be initially most plausible, it would seem that the incompatibility gives us better reason to reject either (i) or (ii). If we are to give up (iii), we need better grounds than its mere incompatibility with (i) and (ii).

Now despite the initial plausibility of (P2'), there are some considerations that could justify us in rejecting this claim. In particular, the plausibility of (P2') will turn in part on some difficult questions about the correct treatment of vagueness. Perhaps we should allow that it can be *vague* whether an agent is in a certain functional state or not. To be clear, what is required is that under an assignment of a functional state to the variable f , it be vague whether αRf holds. Now, it is really far from obvious that this is something that we should countenance. One may, for example, allow that there are certain descriptions or even singular terms that could be substituted for f that would allow for such vagueness without allowing that there could be vagueness under an assignment to the relevant variable. Nonetheless, there might potentially be grounds for allowing for such possibilities. *If* we allowed for such *de re* vagueness, and if we thought that vagueness was a

source of appropriate failures of classical logic, then we would have reason to reject (P2').

Whether one should allow for such cases of *de re* vagueness, and whether one should, in addition, endorse a non-classical treatment of vagueness, are both extremely difficult questions. I won't try to answer these questions here. I'll simply note that whether or not the above argument shows that content atomist views preclude doxastic/credal states from giving rise to failures of classical logic depends in part on how these difficult issues are resolved.

Rejecting (P1)

The other way that our argument can be resisted is by rejecting (P1), i.e., by claiming that $B_\alpha T(\ulcorner \beta \urcorner)$ may exhibit one of the non-classical statuses that, in our models, are reserved for formulas involving conditionals. It's hard to know how seriously to take this response. In order for this to, I think, be a reasonable option we would need to show how to construct VF^+ and SV^+ type models in which non-extremal values that aren't value $1/2$ are assigned to formulas that don't involve conditionals. Our current method for constructing KF^+ and SV^+ models doesn't allow for this possibility and nor, so far as I can see, is there any obvious minor tweak that would allow for this possibility. But of course, this doesn't mean that such models couldn't be constructed, and insofar as this remains a live possibility, we must leave open the option of blocking our argument by rejecting (P1).

A Second Argument

At this point, it is worth noting that in the case of the paracomplete theory a stronger argument can be offered for the claim that a content atomist view precludes the appropriate failures of classical logic in the doxastic/credal realm. This alternative argument also relies on (P2') and so may be blocked if one thinks that vagueness can be a source of appropriate failures of classical logic.

We let M be a VF^+ model. We assume:

- (i) $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket \neq 1$ and $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket \neq 0$.
- (ii) For every assignment function g , either $\llbracket \alpha Rf \rrbracket^g = 1$ or $\llbracket \alpha Rf \rrbracket^g = 0$.

We can show that in M the following holds:

$$\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket \neq 1$$

Proof

Assume that $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket = 1$. Then for some g , $\llbracket (B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket^g = 1$. But this holds only if $\llbracket B_\alpha T(\ulcorner \beta \urcorner) \rrbracket^g = \llbracket \alpha Rf \rrbracket^g$. But, by (i) and (ii), we know that this doesn't hold. So $\llbracket \exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf) \rrbracket \neq 1$.

This argument is in some ways more powerful, and in other ways weaker, than our earlier argument.

It is more powerful in that it relies on weaker assumptions. We have no need to appeal to (P1). Of course, we still need to appeal to (P2'), and so this argument is still hostage to how we come down on the question of the correct treatment of vagueness. Nonetheless, given (P2'), what this argument shows is that a paracomplete theorist who endorses content atomism can't allow $B_\alpha T(\ulcorner \beta \urcorner)$ to have any non-classical status.

This argument, however, is weaker in that it doesn't apply to a paraconsistent theory. The reason for this is that a paraconsistent theorist doesn't think that semantic value 1 is the only designated value. That $\exists f(B_\alpha T(\ulcorner \beta \urcorner) \leftrightarrow \alpha Rf)$ can't have semantic value 1, does not preclude the paraconsistent theorist from endorsing this claim.

Nor do I know of any similar proof for SV^+ models. Perhaps there is such a proof. But if there is, it is not a trivial variant of the proof in the case of VF^+ models.

Summarizing, then, it seems that the question of whether commitment to atomistic contents precludes a non-classical solution to the doxastic and credal paradoxes is open. There are plausible views that would rule out such solutions, given commitment to atomistic contents. But these views are not *clearly* correct. It may be that one can endorse atomistic contents and resolve the doxastic and credal paradoxes in the manner that I've suggested; but whether or not this is the case, turns on some difficult questions.

5.1.2 Holistic Contents and Non-Classical Doxastic/Credal States

The situation for accounts that endorse holistic contents is much more clear cut. In fact, we've already seen that a content holist can allow for failures in classical logic for claims about what an agent believes or what an agent's credence is.

In §4.1.2, we showed that if we represent an agent's doxastic state using a set of VF^+ possible worlds, and take it that an agent counts as believing ϕ just in case ϕ is true at every such world, then certain propositions concerning what the agent believes will have semantic value 1/2. For example, if ϕ is a proposition that has semantic value 1/2 at every possible world in the set representing the agent's doxastic state, it follows that the claim that ϕ is true at every world in that set will also have semantic value 1/2. And so, the claim that the agent believes ϕ will have semantic value 1/2.

In §4.2.1, we showed that if we represent an agent's credal state as a probability measure over a space of VF^+ possible worlds, and take it that an agent has credence x in a proposition ϕ just in case the set of worlds in which ϕ is true has measure x , then certain propositions concerning the agent's credence will have semantic value 1/2. For example, if ϕ is a proposition such that at every possible world ϕ has value 1/2, there will be sets of measure x and sets not of measure x such that for each set the claim that it is the set of worlds in which ϕ is true will have value 1/2. In this case, the claim that the agent has credence x in ϕ will have value 1/2.

The key move here is the use of VF^+ possible worlds in the representation of an agent's doxastic/credal state instead of classical possible worlds. What these proofs assure us of is that if a paracomplete or paraconsistent theorist represents an agent's doxastic or credal state using possible

worlds as understood by that theory, then it will follow that certain claims about what the agent believes or what the agent's credences are will give rise to the sorts of failures of classical logic that such theorists think arise in the semantic domain. This assures us that there is no incompatibility between endorsing a holistic view of contents and allowing for paracomplete or paraconsistent failures of classical logic in the doxastic and credal domains.

The same conclusion also holds for the supervaluationist.

In §4.1.3, it was shown that if we take an agent's belief state to be characterized by a set of SV^+ possible worlds, and we take it that an agent counts as believing ϕ just in case ϕ is true at every such world, then certain propositions concerning whether the agent believes ϕ will have value $1/2$.

In §4.2.2, it was similarly shown that (at least in certain cases) if we take an agent's credal state to be characterized by a probability measure over the space of SV^+ possible worlds, and take it that an agent has credence x in a proposition ϕ just in case the set of worlds in which ϕ is true has measure x , then certain propositions concerning whether the agent has credence x in ϕ will have value $1/2$.

A supervaluationist who is a content holist can, therefore, allow for there to be the appropriate failures of classical logic for statements concerning what an agent believes, or what the agent's credences are.

Indeed, for all of our non-classical theories, this conclusion follows naturally from thinking about doxastic/credal states in content holistic terms. If one is a non-classical theorist, and one thinks that a doxastic/credal state just is a state that divides up the space of possible worlds, then one should think that such a state can be represented by a set (or a measure over a set) of possible worlds understood in non-classical terms. What our earlier results show us is that it is a very simple and natural consequence of thinking about doxastic states in this way that in certain cases claims about what an agent believes, or what an agent's credences are, will themselves be sources of similar failures of classical logic. Far from being strange or surprising, then, doxastic and credal non-classicality simply falls out of this way of thinking about doxastic states, together with the thesis that possible worlds are non-classical in character.

5.1.3 Rationality and Mental Content

What we've seen is that the plausibility of the claim that there can be doxastic/credal non-classicality depends in part on certain commitments about the nature of doxastic/credal states. On the atomistic view of content, it is, at the very least, an open question whether there can be doxastic or credal non-classicality. On the content holistic view, doxastic and credal non-classicality seems perfectly natural given the existence of non-classicality for other propositions.

Here it is worth mentioning another view about the nature of doxastic/credal states that is potentially relevant to the plausibility of doxastic/credal non-classicality. According to some, principles of rationality are constitutive of doxastic/credal states. Views of this type are advocated, for example, by David Lewis, Robert Stalnaker, and Donald Davidson.⁸ Here are two not implausible claims about the connection between principles of rationality and doxastic/credal states.

⁸See Lewis (1974), Lewis (1999), Stalnaker (1984), Davidson (1980b,a).

RATIONAL POSSIBILITY Barring any logical impossibility, doxastic and credal states should at least in principle be capable of meeting the requirements imposed by rationality.

RATIONAL PREFERENCE *Ceteris paribus*, we should prefer ascriptions of content that maximize an agent's rationality.

Using either of these principles, we can draw on our earlier result, which showed that in the case of Alpha the demands of rationality can be met only if certain principles of classical logic fail for certain claims about what Alpha believes, to provide an argument against atomistic theories of content.

RATIONAL POSSIBILITY tells us that it must always in principle be possible for an agent to have rational doxastic states, at least if there is no logical impossibility to meeting the requirements. Since, by paracomplete/paraconsistent/supervaluationist lights, there is no logical impossibility to the agent meeting the requirements imposed by CONSISTENCY and EVIDENCE, then by the lights of these theories it must be possible for an agent to meet these requirements. But if, as we suspect, such failures of classical logic are not possible given the atomistic view of content, then it must be that this view is false.

Of course this argument has force only if one accepts RATIONAL POSSIBILITY. And whether there is such a tight connection between the possibilities for doxastic and credal states, and the principles of rationality is going to be controversial. Perhaps the proponent of the atomist content will not think it too difficult a bullet to bite to simply reject RATIONAL POSSIBILITY. But, insofar as one is attracted to this type of view, one should be correspondingly skeptical of the atomistic view of content.

One view about the nature of doxastic states that would support RATIONAL PREFERENCE is that advocated by David Lewis. According to Lewis, belief and desire states are implicitly defined by a tacit theory of folk-psychology, and this tacit theory takes such states to conform to various rational principles. Where we have two competitors to be the referent of such terms, the competitor that does a better job satisfying the definitional principles will have the better claim to being the referent. (This is not to say that where we have such implicit definitions, the referents of the implicitly defined terms need to, or even can, satisfy *all* of the relevant stipulations. Where satisfaction is not to be had, satisficing will do.) In the doxastic case what this means is that where we have competitors for the referent of various mental state terms those which do a better job satisfying the principles of rationality, i.e., those principles that serve to implicitly definite the mental state terms, will be better candidates for being the referents of such terms.

Amongst the competitors to be the referent of such mental state terms will be functional (or brain) states that are atomic bearers of content, and functional (or brain) states that are holistic bearers of content. Given the *ceteris paribus* preference for rationalizing referents, we can appeal to our earlier arguments to provide an argument against the atomist view. What we've argued is that, in certain cases, this view plausibly condemns an agent to irrationality, in a way that the holist view does not. And, plausibly, there won't be any cases in which the reverse is true, since the holist view just adds the extra flexibility of allowing for doxastic non-classicality. So, the surprising claim that our earlier arguments point towards is that states that are holistic bearers of content will be more eligible to be the referent of mental state terms.

Again, these arguments are only as strong as the principles RATIONAL POSSIBILITY and RATIONAL PREFERENCE. These principles are certainly not Moorean. One would not be irrational if one didn't accept them. Nonetheless, I'm inclined to find them rather plausible. And so I'm inclined to think that we have at least good, *prima facie*, reason to think that our earlier arguments can be leveraged to provide us with some surprising insights into the nature of mental states. That is, we have good reason to think that such states have holistic contents, and that they give rise to certain failures of classical logic.

5.1.4 Non-classical Credal States and Decision Theory

I've argued that there are plausible views about the nature of doxastic/credal states that would allow us to make sense of the idea that such states may be sources of certain failures of classical logic. Moreover, I've argued that if one thinks that principles of rationality play a constitutive role in determining mental contents, then we have good additional reason to suppose that such states can give rise to such failures of classical logic.

However, there is a natural worry that arises at this point. Doxastic/credal states do not live in a vacuum. Instead, they interact with other mental states in such a way that they cause agents to act in various ways. This is one of the most fundamental facts about doxastic/credal states. If, then, we are to allow that doxastic/credal states may give rise to the sorts of failures of classical logic under consideration, we would first need assurance that this will not preclude such states from playing the role that they do in the explanation of action.

This is certainly a reasonable demand. Luckily, it can be met.

Decision theory provides us with a model of how credal states explain an agent's actions.⁹ We suppose that an agent has, in addition to a credal state C , a state of desire \mathcal{D} . The agent's desire state determines the strength of the agent's preferences amongst various possibilities. We can model \mathcal{D} as a function (unique up to positive linear transformation) that maps propositions to real numbers.

Given an agent with a credal state C and a desire state \mathcal{D} , we can model a **decision situation** as follows. We let \mathbf{S} be a partition of the space of possible worlds \mathbf{W} , i.e., a set of subsets of \mathbf{W} , such that (i) no two members of \mathbf{S} intersect, and (ii) $\bigcup \mathbf{S} = \mathbf{W}$. Intuitively, members of \mathbf{S} are possible states of the world that the agent takes to be out of its control. We let \mathbf{A} also be a partition of the space of possible worlds. Members of \mathbf{A} are propositions that the agent takes to be within its power to bring about. We let \mathbf{O} be the partition of \mathbf{W} whose members are intersections of members of \mathbf{S} and \mathbf{A} . That is we have: $o \in \mathbf{O} \leftrightarrow \exists s \in \mathbf{S} \exists a \in \mathbf{A} o = s \cap a$. In the case in which \mathbf{S} and \mathbf{A} are both finite, we can represent our decision situation in the form of a matrix.

⁹See e.g., Savage (1972), Jeffrey (1983), and Joyce (1999).

	S ₁	S ₂	S ₃	S ₄	S ₅
A ₁					
A ₂					
A ₃					
A ₄					
A ₅					

Here the rows represent the actions that the agent takes to be available, i.e., the members of **A**. The columns represent possible states of the world that the agent takes to be outside of its control, i.e., the members of **S**. The cells represent the outcomes of the various actions in the various states, i.e., the members of **O**.

Standard decision theory tells us that in a decision situation the optimal act is the one with the highest expected utility. The expected utility of an action *A* is calculated as follows¹⁰:

$$EU(A) = \sum_i \mathcal{D}(A \cap S_i) C(S_i)$$

That is, expected utility is the weighted average of the utilities of the outcomes of an act *A*, as determined by the agent's state of desire, where the weightings are determined by the agent's credal state.

Using the decision theoretic machinery we can provide a model of the role that an agent's credences play in the explanation of action. We can explain an agent's action by modeling their deliberative situation as a decision problem. The agent's performance of the action is explained in terms of the action maximizing expected utility.

This type of model of the role that an agent's credences play in the explanation of action is surely limited. We must allow that in certain cases an agent acts in various irrational ways. Nonetheless it seems plausible that this type of story will cover a wide range of cases. If, then, we are worried that the postulation of credal non-classicality may undermine the role that such states play in the explanation of action, we can go a long way towards addressing these worries if we can show that credal non-classicality need not effect our ability to provide an account of expected utility.

But given what has been said so far, it should be obvious that we can provide an account of the credal and doxastic states of an agent that will allow us to provide calculations of the expected utility of actions, while also allowing for credal non-classicality. What we noted earlier was that if we model an agent's credal state in terms of a probability measure over a space of para-complete/paraconsistent/supervaluationist possible worlds, then certain claims about the agent's credences will give rise to certain failures of classical logic. The important point for our purposes here is that even though, on this account, classical logic will fail for certain claims about what the agent's credences are, classical logic will hold concerning what credence the agent has for various subsets of the space of possible worlds **W**.

¹⁰According to *evidential decision theory* the appropriate weightings are given not by the agent's unconditional credences, but by the agent's credences conditional on *A*. In what follows, I'll ignore evidential decision theory, but all of the points I make go through *mutatis mutandis* for this alternative account.

If we take it that an agent's credences should be modeled in this way, and that similarly, an agent's desire state should be modeled by a function mapping sets of paracomplete/paraconsistent/supervaluationist possible worlds to real numbers, then we can adopt the standard decision theoretic machinery wholesale. A decision problem can be represented in terms of (i) a partition of the space of paracomplete/paraconsistent/supervaluationist possible worlds, \mathbf{S} , representing states of the world which the agent takes to be outside of its control, and (ii) a partition of the space of paracomplete/paraconsistent/supervaluationist possible worlds, \mathbf{A} , representing actions available to the agent. The expected utility of the members of \mathbf{A} can then be calculated in the manner specified above, given the agent's credal state C and desire state \mathcal{D} . The only difference between this model and standard decision theoretic models is that the points over which C and \mathcal{D} are defined are to be understood in paracomplete/paraconsistent/supervaluationist terms. But this won't have any effect on the ability to calculate expected utility.¹¹

Any account of doxastic/credal states must be able to explain the role that such states play in the explanation of action. An account that posits doxastic/credal non-classicality has available at least one plausible model of how states that may give rise to certain failures of classical logic may nonetheless play a certain characteristic role in the explanation of action.

5.2 Non-Classical Epistemic States

So far we've considered whether or not there are reasons (beyond a general rejection of non-classicality) to resist the idea that an agent's doxastic or credal states may be a source of certain failures of classical logic. The one worry that we noted was the following: There are some, not implausible, accounts that hold that, for any proposition ϕ , the truth of the claim that the agent believes ϕ will align with the truth of the claim that the agent is in some functional or neural state. But plausibly such functional or neural states don't give rise to appropriate failures of classical logic, and so whether or not an agent believes ϕ can't itself be a source of such failures of classical logic. We noted ways in which this argument could be resisted. One way would be to deny that there is an appropriate mapping from states of believing ϕ to functional/neural states, but only from total doxastic states to such functional/neural states. Another would be to allow for appropriate non-classicality with respect to the characterizations of an agent's functional or neural states.

At this point it's worth noting that in the case of epistemic states, this type of worry would seem to be *much* less pressing. Certainly we can't take such states to be identical to neural or functional states of the agent.

¹¹It's worth noting that this argument makes two crucial assumptions. It assumes, first, that the states of the world \mathbf{S} that are relevant to a decision situation form a partition of \mathbf{W} , and second, that the options available to the agent \mathbf{A} that are relevant to a decision situation form a partition of \mathbf{W} . I think that these are pretty plausible assumptions but perhaps when we move to a paracomplete/paraconsistent/supervaluationist theory there might be reasons to question these assumptions. Perhaps. But for my purposes this doesn't really matter. For the point that I want to make here is just that credal non-classicality is not something that *in itself* undermines our ability to explicate the role that such states play in the explanation of action by appeal to standard decision theoretic machinery. And the argument just presented is enough to establish this.

Indeed, in contrast to the doxastic/credal case, it would seem that if there were some reductive account to be offered about the nature of $K\phi$, we may be able to use such an account to provide an argument that $K\phi$ can give rise to the appropriate failures of classical logic. To see how this works, note first that any such account would tell us that $K\phi$ is some state X plus the truth of ϕ . Of course, we might be skeptical that any such analysis could be provided, but for now let's bracket such concerns.

Assuming such an analysis, we can provide the following argument for the claim that *if* one allows that certain propositions give rise to either paracomplete or paraconsistent failures of classical logic, then so too will certain claims about whether an agent knows ϕ .

Take an arbitrary VF^+ model. Assume that $\llbracket X \rrbracket \neq 0$ and that $\llbracket \phi \rrbracket \neq 1$ and $\llbracket \phi \rrbracket \neq 0$. Given these assumptions we'll have $\llbracket X \wedge \phi \rrbracket \neq 1$ and $\llbracket X \wedge \phi \rrbracket \neq 0$. So assuming that it is possible for there to be a case in which ϕ is source of failures of classical logic, but the additional factors X that contribute to knowledge that ϕ are such that either they hold or they are themselves a source of a paracomplete/paraconsistent failure of classical logic, we will have a case in which the claim that the agent knows that ϕ is a source of either a paracomplete/paraconsistent failure of classical logic.

The assumption would seem to hold, e.g., on the analysis where X is justified belief that ϕ . Let ϕ , e.g., be the proposition expressed by the liar sentence. It is, I think, plausible to think that one could believe this proposition. It is also, I think, plausible that one could be such that if one believes the proposition then one's belief is justified. For example, one could have formed the belief on the basis of extremely reliable testimony.

Now there are some views, indeed views to which I am attracted, according to which we can't simply assume that it is possible for an agent to believe the proposition expressed by the liar sentence. For example, if one is a paracomplete theorist and one thinks that an agent's doxastic state should be characterized in terms of a set of paracomplete possible worlds, then there will be no possible doxastic state in which the agent determinately believes this proposition. But even in this case our assumption can still hold. For in this case the claim that the agent believes that propositions will itself be a source of failures of classical logic. Moreover, it would seem possible for an agent to be such that it only forms beliefs when those beliefs are justified by the agents evidence. In this case, then, the claim that the agent's belief is justified would have to itself be a source of a paracomplete failure of classical logic.

This argument can't immediately be extended to the supervaluationist case. The reason is as follows. The above argument relied on the following general claim about VF^+ models: for any two propositions ϕ and ψ , if $\llbracket \phi \rrbracket^{vf^+} \neq 0$ and $\llbracket \psi \rrbracket^{vf^+} \neq 1$ and $\llbracket \psi \rrbracket^{vf^+} \neq 0$, then $\llbracket \phi \wedge \psi \rrbracket^{vf^+} \neq 1$ and $\llbracket \phi \wedge \psi \rrbracket^{vf^+} \neq 0$. But this claim does not hold in general in the supervaluationist case. For example let $\psi = \neg\phi$. Nonetheless, I don't think that we should take too seriously the idea that, according to the correct supervaluationist account, for any proposition ϕ that may give rise to supervaluationist failures of classical logic, $X \wedge \phi$ must have semantic value 0, where X is whatever is required for knowledge beyond truth. This claim strikes me as rather implausible. In general, the inference to which we appealed in the VF^+ case will fail according to a supervaluationist just in case there is some analytic connection between ϕ and ψ that guarantees that the conjunction must be false. It would be rather incredible if X , whatever it is, was such that analytically it was guaranteed that it couldn't obtain at the same time as, say, the liar proposition. So, despite the absence of

nearly as clear an argument in the supervaluationist case, I'm inclined to think that, in the presence of some analysis of knowledge, we could argue that given that certain propositions provide for supervaluationist failures of classical logic, there can also be such failures arising from claims about an agent's knowledge.

Of course, this consideration relies on the rather contentious claim that there is some informative analysis of knowledge that could be provided that would factor it into a truth component and a remaining component X . Perhaps there is no such analysis to be had. Perhaps. But insofar as it is at present an open possibility, we should leave it open that non-classicality with respect to a proposition ϕ can give rise to non-classicality with respect to $K\phi$.

It's worth noting another way in which the case for the non-classicality of knowledge would seem to be simpler than the case for the non-classicality of belief or credence.

The paradox of belief, developed in §3.1, and the paradox of credence, developed in §3.2, involved *normative* constraints on belief and credence. Now these paradoxes can, I've argued, be used to argue for the possibility of doxastic and credal non-classicality. This argument, however, requires appeal to the idea that we should take the principles of rationality to be a guide to the nature of doxastic/credal states. Now, this is, I think, an attractive idea, but it is also controversial.

In the case of the paradox of knowledge, developed in §3.3, however, we don't appeal to *normative* constraints. Instead the paradox relies simply on some very plausible *descriptive* principles about knowledge, viz., that knowledge is factive, and that if one knowingly deduces ϕ by known valid means then one knows ϕ . And both of these principles are really extremely plausible. It would seem, I think, not at all misguided to presume that these principles can serve as guides to the nature of knowledge. Of course, we may ultimately want to give up this presumption. For example, this would seem to be the right course if one were antecedently committed to the truth of classical logic. But if one allows for failures of classical logic, then it seems to me that we should maintain this presumption and conclude that knowledge is another potential source of failures of classical logic.

Chapter 6

Non-Classical Cognitive Significance

In the preceding chapters, I've argued that if one endorses a paracomplete/paraconsistent/supervaluationist approach to the semantic paradoxes, then one should hold that in certain cases a rational agent will be such that claims about whether the agent believes certain propositions (or what the agent's credences are in certain propositions) will give rise to the same sorts of failures of classical logic as arise in cases of semantic paradox. What I now want to do is show how this claim can be used to shine some light on a difficult question about how we should understand these non-classical approaches to the semantic paradoxes.

Each of the non-classical approaches to the semantic paradoxes we've considered posits a number of semantic statuses that aren't present in the classical case. If we want to understand what these theories are telling us, part of what we need to settle is the question of what *cognitive significance* these non-classical statuses have. Put another way, we want to know: if one believes that a proposition ϕ has such a non-classical status, what rational constraints does this impose on the attitude that one has towards ϕ ? Or, more generally: if one has a certain credence in the claim that ϕ has some non-classical status, what rational constraints does this impose on the credence one has in ϕ ?

If two accounts agree about the characterization of the validity relation, but disagree about the cognitive significance of certain non-classical statuses employed in the characterization of that relation, then, in an important sense, these are distinct and incompatible accounts of the semantic paradoxes. Insofar, then, as we haven't provided an account of the cognitive significance of the non-classical statuses posited by the paracomplete/paraconsistent/supervaluationist theories, we don't really understand what these non-classical accounts are telling us about paradoxical propositions such as that expressed by the liar sentence.

Proponents of paracomplete, paraconsistent and supervaluationist theories have typically been aware of the need to give an account of the cognitive significance of non-classical statuses employed by the theory. In fact, there are pretty standard accounts of the cognitive significance of these non-classical statuses. In this chapter, I'll show how our earlier results can be used to provide an argument that the standard answers about the cognitive significance of the non-classical statuses are wrong. I'll then outline alternative answers for each account. What results from these considerations is a very different understanding of the significance of our non-classical treatments

of the semantic paradoxes.

I'll begin by looking at the paracomplete case and then consider how things stand with super-valuationist and paraconsistent theories.

6.1 Paracomplete Cognitive Significance

The paracomplete theory that we've been working with posits a great number of non-classical statuses that paradoxical propositions might have. These can be expressed by various iterations of the indeterminacy operator I .¹ What we want to know is: what is the cognitive significance of these non-classical statuses?

On the way to providing a general answer to this question, let's focus for now on a simpler question. Consider the non-classical status expressed by a single occurrence of the indeterminacy operator. What is the cognitive significance of this status? In trying to answer this question, I'll first focus on qualitative attitudes, i.e., belief, disbelief, agnosticism. Later I'll consider the cognitive significance of this status for credences.

6.1.1 Paracomplete Cognitive Significance: Qualitative Attitudes

Letting ϕ be some proposition that one ought to believe is indeterminate, the question we first want to answer is the following: what attitude should one have towards ϕ ? This is what we earlier called *the attitudinal question*.

The orthodox answer to this question is:

REJECTION For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is indeterminate, that one ought to reject both ϕ and its negation.²

To say that this answer is orthodox is in some ways to undersell how wide is the agreement on this point. To my knowledge, *all* prominent defenders of a paracomplete theory have either explicitly or implicitly endorsed the view that rejection is the correct attitude to take towards the proposition expressed by the liar sentence.³

In what follows, I'll argue that if we're paracomplete theorists we should reject REJECTION. Instead, I claim that the correct answer to the attitudinal question is:

INDETERMINACY For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is indeterminate, that one ought to be such that it is indeterminate whether one believes ϕ .⁴

¹Or more generally by various complex operators formed by negation and the determinacy operator D .

²Read: $OBI\phi \models OR\phi \wedge OR\neg\phi$. Rejection of a proposition ϕ can be thought of as having an appropriately low credence in ϕ . Note that for the proponent of REJECTION, rejection of ϕ does not require that one have a high credence in $\neg\phi$.

³See Field (2008) and Field (2003) for this sort of view. Parsons (1984) is the first explicit endorsement that I know of the rejectionist line. The view can, however, be seen as being implicit in parts of Kripke (1975). See also Soames (1999) and Richard (2008).

⁴Read: $OBI\phi \models OIB\phi$.

For rational agents, indeterminacy in the objects of their doxastic states will filter up to the doxastic states themselves.

First, let's look at the negative argument against REJECTION

Rejecting REJECTION

In §4.1.2, we saw that a paracomplete theorist can block the argument given in §3.1 that purported to show the incompatibility of CONSISTENCY, EVIDENCE and POSSIBILITY. If one is a paracomplete theorist and one accepts CONSISTENCY and EVIDENCE (as, I think, one ought to) then one will hold that if Alpha is rational it will be such that excluded-middle fails for the claim that it believes that $\ulcorner\beta\urcorner$ is true. Having arrived at this conclusion, we can provide an argument against REJECTION. What I'll argue is that although a paracomplete theorist can resolve the normative paradoxes developed in §3.1, doing so requires that one reject REJECTION.

The argument for this claim requires a further premiss. What we have seen is that meeting the requirements imposed by CONSISTENCY and EVIDENCE demands that excluded-middle fail for the claim that Alpha believes that $\ulcorner\beta\urcorner$ is true. Now *one way* for this to be the case is that it be *indeterminate* whether Alpha believes that $\ulcorner\beta\urcorner$ is true (indeed, this is the status of Alpha's doxastic state concerning $\ulcorner\beta\urcorner$ in our VF^+ models). But as we've already noted this is not the only status that is compatible with the failure of excluded-middle. An adequate treatment of the semantic paradoxes that appeals to indeterminacy requires that the indeterminacy operator iterate in a non-trivial manner; we must not, for example, have $II\phi \models I\phi$. In particular, this is required in order to adequately treat higher-order paradoxical sentences that employ the determinacy operator.⁵ In addition to first-order indeterminacy, then, there is indeterminate indeterminacy, and indeterminate indeterminate indeterminacy etc. Each of these is such that when a proposition has the status in question excluded-middle fails for that proposition. It doesn't *follow*, then, from the fact that Alpha ought to be such that excluded-middle fails for the claim that it believes that $\ulcorner\beta\urcorner$ is true, that Alpha ought to be such that it is indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true. Luckily we don't need such a strong claim to mount an argument against REJECTION. All that we need is the claim that a *rational* way for Alpha to meet the requirement that it be such that excluded-middle fail for the proposition that it believes that $\ulcorner\beta\urcorner$ is true is for it to be *indeterminate* whether it believes that $\ulcorner\beta\urcorner$ is true.

Later I'll consider how to reframe the argument against REJECTION if we reject the claim

⁵To see this, consider, for example, the sentence λ^* which is provably equivalent to $\neg DT\ulcorner\lambda^*\urcorner$. Given excluded-middle we can derive a contradiction using λ^* . First, assume λ^* . In general $\phi \models D\phi$, and so in particular $\lambda^* \models D\lambda^*$. We have, then, $D\lambda^*$. Given the intersubstitutivity of ϕ and $T\ulcorner\phi\urcorner$ we also have $DT\ulcorner\lambda^*\urcorner$. But it also follows from λ^* that $\neg DT\ulcorner\lambda^*\urcorner$. So a contradiction can be derived from λ^* . Now assume $\neg\lambda^*$. This entails $DT\ulcorner\lambda^*\urcorner$, which entails $T\ulcorner\lambda^*\urcorner$, which in turn entails λ^* . Again we have a contradiction. Assuming that we have $\lambda^* \vee \neg\lambda^*$ we can derive a contradiction outright.

If one wants to treat the liar sentence by rejecting excluded-middle one should extend the same treatment to this case as well. But note that in this case we cannot characterize λ^* as being indeterminate. For this entails $\neg D\lambda^*$ which is provably equivalent to, and so entails, λ^* , which of course lands us right back in paradox. If we want a way of characterizing this sentence's paradoxical status the indeterminacy operator must iterate in a non-trivial manner. We can characterize λ^* as being indeterminately indeterminate; but only if it's not the case that $II\phi \models I\phi$.

that it is rational for Alpha to be such that it is indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true. But I think that it's very hard to see what sort of principled reason there could be for denying this. Given that Alpha is required to be such that excluded-middle fails for the claim that it believes that $\ulcorner\beta\urcorner$ is true, if we are to avoid falling back into normative paradox there should be some way for Alpha to meet this requirement without incurring rational criticism. The question is, then, what are the rational ways for Alpha to meet this requirement, if it being indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true isn't one of them? Is it being indeterminately indeterminate whether Alpha believes that $\ulcorner\beta\urcorner$ is true a rational way to meet the relevant requirement, or it being indeterminately indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true, or some other status? It is very hard to see why some such higher-order status(es) should be rational, while it is irrational for it to be simply indeterminate. Perhaps there is some deep surprising explanation for this. But at present I can't see what that would be, and so I'll proceed for now on the assumption that Alpha may rationally meet the requirements imposed by CONSISTENCY and EVIDENCE by being such that it is indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true.

Given this assumption we can now argue as follows. We can show that although the paracomplete theorist can resolve our earlier normative paradox, given acceptance of REJECTION we can resurrect this paradox in a way that is not amenable to similar solution.

We have assumed that it is indeterminate whether Alpha believes that $\ulcorner\beta\urcorner$ is true, and that this is a rational way for Alpha to meet the obligations imposed by CONSISTENCY and EVIDENCE. This is logically equivalent to the claim that it is indeterminate whether Alpha does not believe that $\ulcorner\beta\urcorner$ is true, that is $I\neg B_\alpha T(\ulcorner\beta\urcorner)$. Let us add to our story about Agent Alpha. We are now allowing Alpha's doxastic states to be indeterminate. We should, therefore, extend our transparency assumptions to take account of this possibility:

$$(18) I\neg B_\alpha T(\ulcorner\beta\urcorner) \leftrightarrow B_\alpha I\neg B_\alpha T(\ulcorner\beta\urcorner)$$

(18) holds in the VF^+ models in which we have represented Alpha's doxastic state. The assumption, then, that an agent with indeterminate doxastic states may at least in principle satisfy this condition is, therefore, reasonable.

Given that it is indeterminate whether Alpha does not believe that $\ulcorner\beta\urcorner$ is true, by (18) it follows that Alpha believes this, that is $B_\alpha I\neg B_\alpha T(\ulcorner\beta\urcorner)$. As in the earlier cases, we assume that Alpha is perfectly reliable in this belief. The following is a theorem: $I\neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow IT(\ulcorner\beta\urcorner)$.⁶ As in the earlier cases, we can assume that Alpha believes this on the basis of the same superlative grounds as us. Given these assumptions it follows that Alpha's evidence makes it certain that $IT(\ulcorner\beta\urcorner)$. By EVIDENCE, it follows that $OB_\alpha IT(\ulcorner\beta\urcorner)$. By REJECTION, then, it follows that $OR_\alpha T(\ulcorner\beta\urcorner)$. If one rejects ϕ it follows that one does not believe ϕ . Assuming, then, that Alpha meets the rational requirement imposed on it by REJECTION, we have $\neg B_\alpha T(\ulcorner\beta\urcorner)$.

This, however, lands us back into normative paradox. We need simply rehearse Case 1 from §3.1. By (4) we have $B_\alpha \neg B_\alpha T(\ulcorner\beta\urcorner)$. We also have that $\neg B_\alpha T(\ulcorner\beta\urcorner) \rightarrow T(\ulcorner\beta\urcorner)$ is a theorem, and that it is believed by Alpha on excellent grounds. It follows that Alpha's evidence makes it certain

⁶In general where $\phi \leftrightarrow \psi$ is a theorem so is $I\phi \leftrightarrow I\psi$.

that $T(\ulcorner\beta\urcorner)$. Assuming compliance with the normative demands imposed by EVIDENCE we have $B_\alpha T(\ulcorner\beta\urcorner)$. But this, of course, is impossible since we already have $\neg B_\alpha T(\ulcorner\beta\urcorner)$.

We have derived a contradiction on the assumption that Alpha, an antecedently rational agent, meets all of the requirements imposed by EVIDENCE, CONSISTENCY and REJECTION. Note that no appeal was made to excluded-middle, nor was any use made of *reductio* or other forms of proof which fail given the approach to the liar under consideration. The same moves that were available to us to reconcile the seeming incompatibility of CONSISTENCY and EVIDENCE are not available in this case. If we are to hold on to CONSISTENCY, EVIDENCE and POSSIBILITY by allowing for doxastic states to be indeterminate we must reject REJECTION.

CONSISTENCY, EVIDENCE and POSSIBILITY seem to me individually and jointly much more plausible normative conditions than REJECTION. Faced with the choice between holding on to CONSISTENCY, EVIDENCE and POSSIBILITY and holding on to REJECTION, it seems clear to me that the former course is preferable.

It is far from obvious what answer one should give to the attitudinal question. It is difficult to get an independent grip on this issue. Given this, we should, I think, take seriously an argument that shows how the answer to this question is constrained by our acceptance of other clearer normative conditions. The incompatibility between CONSISTENCY, EVIDENCE, POSSIBILITY and REJECTION provides a good reason to give up REJECTION.

Generalizing the Argument

I've argued that if we are to avoid resurrecting our normative paradox we can't accept REJECTION. Now, it isn't hard to see that the argument I've presented easily generalizes in the following two ways.

Claim 1: If we want to hold on to CONSISTENCY, EVIDENCE and POSSIBILITY, then we should reject any view that endorses $OBI\phi \models O\neg B\phi$.

In arguing that acceptance of REJECTION leads to the resurrection of our normative paradox, I argued that, given CONSISTENCY and EVIDENCE, Alpha was required to believe that it is indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true. But, given this, in order to meet the requirement imposed by REJECTION, Alpha was forced to *not* believe that $\ulcorner\beta\urcorner$ is true. And, as demonstrated by Case 1, this resulted in a failure to meet a requirement imposed by EVIDENCE. The important feature of REJECTION for the argument, then, was that meeting the requirement it imposes demands that one not believe a proposition that one ought to believe is indeterminate. Thus any view that endorses $OBI\phi \models O\neg B\phi$ will be subject to the same argument.

Claim 2: If we want to hold on to CONSISTENCY, EVIDENCE and POSSIBILITY, then we should reject any view that endorses $OBI\phi \models OB\phi$.

The argument against REJECTION showed that Alpha must believe that it is indeterminate whether it believes that $\ulcorner\beta\urcorner$ is true, in order to meet certain requirements imposed by CONSISTENCY and EVIDENCE. Case 2 earlier showed that if Alpha does believe that $\ulcorner\beta\urcorner$ is true then it will either fail to meet a requirement imposed by CONSISTENCY or a requirement imposed by EVIDENCE. If, then, we were to accept that one ought to believe a proposition that one ought to believe is indeterminate, we would be able to resurrect the normative paradox developed in §3.1. Any view that endorses $OBI\phi \models OB\phi$ is, therefore, subject to a variant of the argument against REJECTION.

The generalization of the argument against REJECTION serves to rule out a number of interesting potential answers to the attitudinal question. Perhaps it isn't terribly surprising to find that belief is not the correct response to indeterminacy. However, one might be tempted to think that if rejection isn't the correct attitude towards cases of indeterminacy, then perhaps agnosticism is. Or perhaps one might be tempted by the thought that in response to perceived cases of indeterminacy one should simply opt out of having any normal doxastic attitude towards such a proposition. That is, one might be tempted by the thought that in response to indeterminacy one shouldn't believe, be agnostic about, or reject such a proposition, but have some other attitude incompatible with these. What the argument I've provided shows is that these options are ultimately just as unacceptable as REJECTION.

INDETERMINACY

I've argued that we shouldn't accept REJECTION as providing the correct answer to the attitudinal question. Indeed, if the argument I've given against REJECTION is right, then we shouldn't accept any answer to the attitudinal question that holds that one shouldn't believe a proposition that one ought to believe is indeterminate, nor should we accept any answer that holds that one should believe such a proposition.

What then is the correct answer to the attitudinal question? One option would be to argue that there is no general normative condition connecting the indeterminacy of propositions and our doxastic states concerning those propositions. According to this line, in standard cases of indeterminacy, such as the liar sentence, REJECTION does give us the right story, but in other cases, such as $\ulcorner\beta\urcorner$, another story is appropriate. Indeterminacy would in this way be like contingency.⁷ There is no single attitude one should have towards propositions that one takes to be contingent. Some we should believe, some we should reject, and others we should simply be agnostic about; it depends on what our evidence tells us.

This is a consistent position, but I don't see that it has much to recommend it. How exactly should we restrict REJECTION? To say that we simply restrict it for those cases in which it leads

⁷Contingency is understood as follows: $C\phi \leftrightarrow_{df} \diamond\phi \wedge \diamond\neg\phi$.

to normative paradox seems hopelessly *ad hoc*. But what other principled distinction can we draw between, say, the case in which an agent believes that the proposition expressed by the liar sentence is indeterminate, and the case in which it believes that the proposition expressed by $\lceil\beta\rceil$ is indeterminate? In the case of contingency we can say something about why an agent may believe both that ϕ is contingent and that ψ is contingent and yet rationally take different attitudes towards the two propositions; the agent may, for example, have conclusive evidence that one is true and the other false. In the case of indeterminacy, however, I have no idea what sort of analogous story one could tell that would make REJECTION deliver the correct verdict in all but the problematic cases.

What I'll argue in this section is that we should instead accept:

INDETERMINACY For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is indeterminate, that one ought to be such that it is indeterminate whether one believes ϕ .

The argument has two parts. I'll first argue that INDETERMINACY is independently motivated. I'll then show that, unlike REJECTION, Alpha is able to satisfy the demands imposed by INDETERMINACY in addition to those imposed by CONSISTENCY and EVIDENCE.

First, the argument for independent motivation.

It is very easy to be puzzled about what answer to give to the attitudinal question. For *prima facie* the following three claims are all plausible:

- (19) If one ought to believe that ϕ is indeterminate, it follows that one ought not believe ϕ .
- (20) If one ought to believe that ϕ is indeterminate, it follows that one ought not be agnostic about ϕ .
- (21) If one ought to believe that ϕ is indeterminate, it follows that one ought not reject ϕ .

(19) will, I suspect, strike you as immediately plausible. Consider, for example, a paradigmatic indeterminate proposition such as that expressed by the liar sentence. In this case, belief would certainly *seem* to be an inappropriate attitude.

It would also seem, as (20) maintains, to be inappropriate to be agnostic towards this proposition. After all, agnosticism is the correct attitude to take towards a proposition about which one takes oneself to be ignorant. One who thinks that the proposition expressed by the liar sentence is indeterminate would not, however, seem to think that there is some fact of the matter concerning the truth value of this proposition about which they are ignorant.

(21) may be less immediately compelling, but one can argue for it by appeal to the following principle:

NEGATION One ought to be such that one rejects a proposition ϕ just in case one believes its negation $\neg\phi$.⁸

⁸Read: $O(R\phi \leftrightarrow B\neg\phi)$. This is a principle that a proponent of REJECTION will reject. But it should be conceded that this is *prima facie* quite plausible.

ϕ is indeterminate just in case $\neg\phi$ is indeterminate. If one ought to believe that ϕ is indeterminate, then one ought to believe that $\neg\phi$ is indeterminate. By (19), then, one ought not believe $\neg\phi$. By NEGATION, one ought to be such that if one does not believe $\neg\phi$ then one does not reject ϕ . Given that doxastic obligations are closed under consequence it follows that one ought not reject ϕ . This gives us (21).

The problem is that the following claim is also quite plausible:

- (22) If one ought to believe that ϕ is indeterminate, it follows that one should not fail to have *some* positive doxastic attitude towards ϕ .

To have a positive doxastic attitude towards a proposition ϕ is (as I'm using the term) to either believe ϕ , to be agnostic about ϕ , or to reject ϕ . Of course, in certain situations it may indeed be permissible to fail to have some such positive doxastic attitude towards a proposition. In particular, where one lacks certain relevant concepts it would, indeed, I think, be wrong to say that rationality demands that the agent form opinions that require the deployment of such concepts. But, given this, if one is required to believe that ϕ is indeterminate, then one will have all the relevant concepts required for forming judgments as to whether or not ϕ obtains. But, given that the agent is equipped with the relevant concepts, it is at least *prima facie* plausible that an agent should have some doxastic attitude towards ϕ .

A paracomplete theorist will, presumably, think that there are cases in which an antecedently rational agent ought to believe that a certain proposition is indeterminate. But, given this and (19)-(22), this agent will be saddled with a set of obligations that it is impossible to meet. (19)-(22), then, are incompatible with POSSIBILITY. What (19)-(22) amount to is the claim that there is simply no rational response to cases of perceived indeterminacy. This is something that, I think, a paracomplete theorist should clearly not accept.

Call the *prima facie* plausibility of (19)-(22) *the normative problem*. An adequate response to the the normative problem should identify which of (19)-(22) we should give up, and *in addition* it should provide a plausible error-theory that can account for the *prima facie* plausibility of (19)-(22).

A proponent of REJECTION can provide the following response to the the normative problem. Such a proponent will say that we should give up (21) but hold on to (19), (20) and (22). To this end the proponent of REJECTION will reject NEGATION. In support of this response she could offer the following plausible error-theory. First, it is worth noting that while NEGATION does not hold unrestrictedly it does hold in those cases in which excluded-middle holds. But then it would seem quite reasonable that we could mistakenly find (21) plausible. For, of course, we are naturally prone to overgeneralize from those cases in which classical logic holds. Thus while (21) is not correct, we can account for its *prima facie* plausibility.

At first glance, this is an attractive response to the normative problem. However, what we have now seen is that it is ultimately unacceptable. For accepting (19), (20) and (22), while giving up (21), commits one to the acceptance of REJECTION. But this, we've seen, saddles us with normative paradox and so is ultimately unacceptable.

Indeed, the generalizations of our argument against REJECTION show that one can't solve the normative problem by simply rejecting one of (19)-(22). If one rejects (22) but holds on to

(19)-(21), then one is committed to the view that if one ought to believe that a certain proposition ϕ is indeterminate, then one ought to not believe ϕ . Similarly if one rejects (20) but holds on to (19), (21) and (22). And if one rejects (19) but holds on to (20)-(22), then one is committed to the view that if one ought to believe that a certain proposition ϕ is indeterminate, then one ought to believe ϕ . Any of these minimal responses to the normative problem, then, leads to an ultimately unacceptable view.

What I will now do is outline an alternative response to the normative problem that is able to do justice to the intuitions that motivate (19)-(22), while avoiding the untoward consequences of accepting these claims. I'll then show how INDETERMINACY is a consequence of this response. Given the paucity of reasonable responses to the normative problem, that INDETERMINACY follows from an elegant error-theoretic response gives us a reason to take INDETERMINACY seriously as the answer to the attitudinal question.

The response to the normative problem that I advocate involves rejecting each of (19)-(22). In their stead, we should accept the following closely related principles:

- (19^d) If one ought to believe that ϕ is indeterminate, it follows that one ought not *determinately* believe ϕ .
- (20^d) If one ought to believe that ϕ is indeterminate, it follows that one ought not be *determinately* agnostic about ϕ .
- (21^d) If one ought to believe that ϕ is indeterminate, it follows that one ought not *determinately* reject ϕ .
- (22^d) If one ought to believe that ϕ is indeterminate, it follows that one ought not *determinately* fail to have some positive doxastic attitude towards ϕ .

Unlike with (19)-(22), an agent can meet each of the requirements imposed by (19^d)-(22^d). And a proponent of (19^d)-(22^d) can provide the following simple error-theory to account for the *prima facie* plausibility of (19)-(22). We are not terribly good at distinguishing between something being the case and its *determinately* being the case. Indeed, insofar as we are able to make this distinction, it is only as the result of significant theoretical work; that there is a distinction only becomes clear when we see how it is necessary in order to resolve certain paradoxes, such as that raised by the liar sentence. We should not expect, then, that in advance of this work our intuitions should be finely attuned to this distinction. If, then, there are true principles, such as (19^d)-(22^d), that concern certain conditions obtaining determinately, it should not be unexpected that we would confuse such principles for other principles, such as (19)-(22), that concern those conditions simply obtaining whether determinately or not. By accepting (19^d)-(22^d), then, we can account for the plausibility of (19)-(22), while avoiding their undesirable consequences.

Next I'll show that INDETERMINACY is a consequence of (19^d)-(22^d); that is I'll show that from (19^d)-(22^d) it follows that $OBI\phi \models OIB\phi$. To show that this is so, it will suffice to show that

both (i) $OBI\phi \models O\neg DB\phi$ and (ii) $OBI\phi \models O\neg D\neg B\phi$ follow from (19^d)-(22^d).⁹

Now, (i) just is (19^d) and so it trivially follows from (19^d)-(22^d).

To show that (ii) is a consequence of (19^d)-(22^d) we can argue as follows. We want to establish $OBI\phi \models O\neg D\neg B\phi$. One fails to believe ϕ just in case one is either agnostic about ϕ or one rejects ϕ or one simply has no positive doxastic attitude towards ϕ . That is we have $\models \neg B\phi \leftrightarrow (A\phi \vee R\phi \vee \neg P\phi)$. So what we want to establish is $OBI\phi \models O\neg D(A\phi \vee R\phi \vee \neg P\phi)$. (20^d) tells us that $OBI\phi \models O\neg DA\phi$. (21^d) tells us that $OBI\phi \models O\neg DR\phi$. And (22^d) tells us that $OBI\phi \models O\neg D\neg P\phi$. Since, I take it, obligations are close under consequence we have $OBI\phi \models O(\neg DA\phi \wedge \neg DR\phi \wedge \neg D\neg P\phi)$. Now, in general we have:

$$(23) \quad \neg D\phi \wedge \neg D\psi \wedge \neg D\xi \models \neg D(\phi \vee \psi \vee \xi).^{10}$$

So in particular we have $\neg DA\phi \wedge \neg DR\phi \wedge \neg D\neg P\phi \models \neg D(A\phi \vee R\phi \vee \neg P\phi)$. And so given the closure condition on rational obligations we have $OBI\phi \models O\neg D(A\phi \vee R\phi \vee \neg P\phi)$. This suffices to establish (ii).

Since (i) and (ii) follow from (19^d)-(22^d), it follows that INDETERMINACY is a consequence of (19^d)-(22^d). The latter, I've argued, provide an attractive response to the normative problem. This is a reason to think that INDETERMINACY gives the correct answer to the attitudinal question.

Having argued that INDETERMINACY is independently motivated, the next point to make is that, unlike REJECTION, Alpha can satisfy the demands imposed by INDETERMINACY while satisfying the demands imposed by CONSISTENCY and EVIDENCE.

⁹To see that this will suffice, note that, since $IB\phi$ is equivalent to $\neg DB\phi \wedge \neg D\neg B\phi$, it follows that $OBI\phi \models OIB\phi$ is equivalent to $OBI\phi \models O(\neg DB\phi \wedge \neg D\neg B\phi)$. And as an instance of a general closure principle we have: $O\neg DB\phi, O\neg D\neg B\phi \models O(\neg DB\phi \wedge \neg D\neg B\phi)$. Thus if we can show that (i) and (ii) hold then we can show that $OBI\phi \models O(\neg DB\phi \wedge \neg D\neg B\phi)$ and so $OBI\phi \models OIB\phi$.

¹⁰Instead of proving (23), I'll sketch the proof for the simpler $\neg D\phi \wedge \neg D\psi \models \neg D(\phi \vee \psi)$. The proof of (23) is a straightforward generalization of this. The cases one needs to deal with are greater, but add no further complexities.

Let $\llbracket \neg D\phi \wedge \neg D\psi \rrbracket = 1$. We want to show $\llbracket \neg D(\phi \vee \psi) \rrbracket = 1$. By a general theorem that Field calls the FUNDAMENTAL THEOREM, we know that at arbitrarily high stages in the construction of a VF^+ model, there will be ordinals such that a sentence ϕ will have value 1 at that ordinal just in case it has value 1 in the VF^+ model, and similarly a sentence ϕ will have value 0 at that ordinal just in case ϕ has value 0 in the VF^+ model. See Field (2007) p.257-258 for a proof of this claim. Let's denote the least such ordinal σ . To show that $\llbracket \neg D(\phi \vee \psi) \rrbracket = 1$ follows on the assumption that $\llbracket \neg D\phi \wedge \neg D\psi \rrbracket = 1$, it suffices to show that $\llbracket \neg D(\phi \vee \psi) \rrbracket_\sigma = 1$ follows on the assumption that $\llbracket \neg D\phi \wedge \neg D\psi \rrbracket_\sigma = 1$.

Now, $\llbracket \neg D\phi \wedge \neg D\psi \rrbracket_\sigma = 1$ holds just in case $\llbracket \neg(\phi \wedge \neg(\phi \rightarrow \neg\phi)) \wedge \neg(\psi \wedge \neg(\psi \rightarrow \neg\psi)) \rrbracket_\sigma = 1$. And this holds just in case either (i) $\llbracket \phi \rrbracket_\sigma = \llbracket \psi \rrbracket_\sigma = 0$ or (ii) $\llbracket \phi \rrbracket_\sigma = 0$ and $\llbracket \psi \rightarrow \neg\psi \rrbracket_\sigma = 1$, or (iii) $\llbracket \psi \rrbracket_\sigma = 0$ and $\llbracket \phi \rightarrow \neg\phi \rrbracket_\sigma = 1$, or (iv) $\llbracket \phi \rightarrow \neg\phi \rrbracket_\sigma = \llbracket \psi \rightarrow \neg\psi \rrbracket_\sigma = 1$.

$\llbracket \neg D(\phi \vee \psi) \rrbracket_\sigma = 1$ just in case $\llbracket (\phi \vee \psi) \wedge \neg((\phi \vee \psi) \rightarrow \neg(\phi \vee \psi)) \rrbracket_\sigma = 0$. We'll show that the latter holds whichever of (i)-(iv) obtains.

If (i) holds then $\llbracket \phi \vee \psi \rrbracket_\sigma = 0$, and so clearly $\llbracket (\phi \vee \psi) \wedge \neg((\phi \vee \psi) \rightarrow \neg(\phi \vee \psi)) \rrbracket_\sigma = 0$.

If (ii) holds then we have that there exists a $\beta < \sigma$ such that for all γ such that $\beta \leq \gamma < \sigma$ $\llbracket \psi \rrbracket_\gamma \leq 1/2$. Since σ is an acceptable ordinal we can be assured that given that $\llbracket \phi \rrbracket_\sigma = 0$ it follows that for all γ such that $\beta \leq \gamma < \sigma$ $\llbracket \phi \rrbracket_\gamma = 0$. Thus for all γ such that $\beta \leq \gamma < \sigma$ $\llbracket \phi \vee \psi \rrbracket_\gamma \leq 1/2$. This assures us that $\llbracket ((\phi \vee \psi) \rightarrow \neg(\phi \vee \psi)) \rrbracket_\sigma = 1$, and so $\llbracket (\phi \vee \psi) \wedge \neg((\phi \vee \psi) \rightarrow \neg(\phi \vee \psi)) \rrbracket_\sigma = 0$.

Proving the cases for (iii) and (iv) simply involves applying the reasoning from the previous two cases in an obvious way.

Alpha is an agent who believes the theorem $T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$ and is, in addition, doxastically self-transparent with respect to the proposition that it believes that $\ulcorner\beta\urcorner$ is true. I noted that we could represent the doxastic state of an agent satisfying these stipulations by a paracomplete possible-worlds model in which the accessibility relation is an equivalence relation. In such a model, the agent's beliefs will be consistent and closed under logical consequence. This assured us that an agent such as Alpha could in principle meet the demands imposed by CONSISTENCY and EVIDENCE. In this model, however, it is indeterminate whether Alpha believes that $\ulcorner\beta\urcorner$ is true.

This model is also sufficient to assure us that Alpha is able to meet whatever additional demands might be imposed by INDETERMINACY. First note that, in general, for any class of VF^+ models \mathcal{M} , such that for every point δ there is some R -accessible δ' , the following holds:

$$(24) B_\alpha I\phi \models_{\mathcal{M}} IB_\alpha\phi$$

Proof of (24)

Assume that $\llbracket B_\alpha I\phi \rrbracket^\delta = 1$. We'll show that $\llbracket IB_\alpha\phi \rrbracket^\delta = 1$. To do this it will suffice to show (i) for any δ , $\llbracket I\phi \rrbracket^\delta = 1$ just in case $\llbracket \phi \rrbracket^\delta = 1/2$. For, given (i), we can argue as follows. From $\llbracket B_\alpha I\phi \rrbracket^\delta = 1$ it follows that for all δ' such that $\delta R\delta'$, $\llbracket I\phi \rrbracket^{\delta'} = 1$. And, given (i), it follows that $\llbracket \phi \rrbracket^{\delta'} = 1/2$. And, so we have $\llbracket B_\alpha\phi \rrbracket^\delta = 1/2$, and thus $\llbracket IB_\alpha\phi \rrbracket^\delta = 1$.

To show that (i) holds, we need to show that (ia) $\llbracket I\phi \rrbracket^\delta = 1 \Rightarrow \llbracket \phi \rrbracket^\delta = 1/2$, and (ib) $\llbracket I\phi \rrbracket^\delta = 1 \Leftarrow \llbracket \phi \rrbracket^\delta = 1/2$. (ib) is trivial, and I leave this to the interested reader to verify. To show that (ia) holds, we will assume that $\llbracket I\phi \rrbracket^\delta = 1$ and that at some stage γ in our VF^+ model, either $\llbracket \phi \rrbracket_\gamma^\delta = 1$ or $\llbracket \phi \rrbracket_\gamma^\delta = 0$. We'll show that this package of assumptions leads to a contradiction. This suffices to establish (ia).

Note that $I\phi$ is equivalent to $\neg(\phi \wedge \neg(\phi \rightarrow \neg\phi)) \wedge \neg(\neg\phi \wedge \neg(\neg\phi \rightarrow \phi))$. So the assumption, then, that $\llbracket I\phi \rrbracket^\delta = 1$ is equivalent to the assumption that $\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket^\delta = \llbracket \neg\phi \wedge \neg(\neg\phi \rightarrow \phi) \rrbracket^\delta = 0$.

What we want to show is that if at any stage γ of our VF^+ model either $\llbracket \phi \rrbracket_\gamma^\delta = 1$ or $\llbracket \phi \rrbracket^\delta = 0$, then there is some stage α such that $\llbracket I\phi \rrbracket_\alpha^\delta \neq 1$. We argue for this as follows. First we show that, given that $\llbracket I\phi \rrbracket^\delta = 1$, it follows that if $\llbracket \phi \rrbracket_\gamma^\delta = 1$ then $\llbracket \phi \rrbracket_{\gamma+1}^\delta = 0$, and similarly, if $\llbracket \phi \rrbracket_\gamma^\delta = 0$ then $\llbracket \phi \rrbracket_{\gamma+1}^\delta = 1$. To see this note that if $\llbracket \phi \rrbracket_\gamma^\delta = 1$, then $\llbracket \neg\phi \rightarrow \phi \rrbracket_{\gamma+1}^\delta = 1$. This means that $\llbracket \neg\phi \wedge \neg(\neg\phi \rightarrow \phi) \rrbracket_{\gamma+1}^\delta = 0$. But, then, since $\llbracket \phi \rightarrow \neg\phi \rrbracket_{\gamma+1}^\delta = 0$ it is clear that in order for $\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket_{\gamma+1}^\delta = 0$ we must have $\llbracket \phi \rrbracket_{\gamma+1}^\delta = 0$. A similar argument shows that, given that $\llbracket I\phi \rrbracket^\delta = 1$, if $\llbracket \phi \rrbracket_\gamma^\delta = 0$ then $\llbracket \phi \rrbracket_{\gamma+1}^\delta = 1$. So what this shows is that if at some stage γ , $\llbracket \phi \rrbracket_\gamma^\delta = 1$ or $\llbracket \phi \rrbracket^\delta = 0$, then given that $\llbracket I\phi \rrbracket^\delta = 1$, the value of ϕ will alternate between 0 and 1 for the successor ordinals following γ , prior to the next limit ordinal λ . But given this, we can show that at the next limit ordinal λ $\llbracket I\phi \rrbracket_\lambda^\delta \neq 1$. To see this it suffices to note that given the alternating pattern it follows that $\llbracket \phi \rightarrow \neg\phi \rrbracket_\lambda^\delta = \llbracket \neg\phi \rightarrow \phi \rrbracket_\lambda^\delta = 1/2$. And give this, it follows that it can't be the case that $\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket^\delta = \llbracket \neg\phi \wedge \neg(\neg\phi \rightarrow \phi) \rrbracket^\delta = 0$. This suffices, then, to establish that given that $\llbracket I\phi \rrbracket^\delta = 1$, it follows that $\llbracket \phi \rrbracket^\delta = 1/2$.

Given (24), it follows that in any of the models in which Alpha meets the demands imposed by CONSISTENCY and EVIDENCE, Alpha will also meet the demands imposed by the combination of CONSISTENCY, EVIDENCE and INDETERMINACY.

To see this, first note that INDETERMINACY, which states $OB_\alpha I\phi \models OIB_\alpha\phi$, only issues in an obligation given an input of the form $OB_\alpha I\phi$. We can think about the obligations that result from CONSISTENCY, EVIDENCE and INDETERMINACY as the result of the following iterated process. We start with the obligations that result from CONSISTENCY and EVIDENCE. These are the obligations that result from CONSISTENCY, together with the obligations that result from EVIDENCE, together with whatever obligations follow from these given general principles of deontic logic. (In what has preceded, and in what follows, the only general principle of deontic logic I'm assuming is that the logical consequences of a set of rational obligations are themselves rationally obligatory.) Label this set Ω^0 . This delivers possible inputs to INDETERMINACY. This delivers further obligations, which, together with general principles of deontic logic, gives us a set of obligations Ω^1 . And so on. The end result of this process is the set of obligations entailed by CONSISTENCY, EVIDENCE and INDETERMINACY.

Given a model M in which our agent meets CONSISTENCY and EVIDENCE, (24) assures us that at each stage of this process the agent will, in M , satisfy the obligations that result at that stage. By hypothesis the agent in M meets Ω^0 . By (24), in M the agent will meet whatever obligations result from Ω^0 together with INDETERMINACY. In such a model, moreover, if a set of obligations are met, then so is any logical consequence of this set. So the agent will meet Ω^1 . And it is clear that the reasoning here generalizes. For each stage γ , Alpha will will meet the obligations imposed by Ω^γ .

We can be assured, then, given that Alpha is able to meet the obligations that follow from CONSISTENCY and EVIDENCE, that it can meet any additional obligations that result from the endorsement of INDETERMINACY. INDETERMINACY, unlike REJECTION, does not land us back into normative paradox.

Further Reasons for Rejecting REJECTION (and a More General Account of Paracomplete Cognitive Significance)

A key premiss in the preceding argument was that a *rational* way for Alpha to meet the requirements imposed by CONSISTENCY and EVIDENCE was for Alpha to be such that it was indeterminate whether it believed that $\lceil\beta\rceil$ is true. I noted that I couldn't see any principled reason for denying this claim. Nonetheless, it is, I think, worth considering how matters stand if one does not rely on this assumption.

As I noted earlier, not every paradoxical proposition for which excluded-middle fails can be characterized as being indeterminate. Some we can only characterize as being indeterminately indeterminate, others as being indeterminately indeterminately indeterminate, and so on. The attitudinal question that we've been focused on, then, is a member of a tightly knit family of questions: Given that one ought to believe that a proposition ϕ is indeterminateⁿ what attitude should one have towards ϕ ?

Prima facie, it is natural for the proponent of REJECTION to give the same answer to each of

these questions.¹¹ However, I think that the *minimal* lesson that we should draw from the earlier discussion of the normative paradoxes is that if one is a paracomplete theorist then one should not accept that rejection is *always* the correct attitude to take towards propositions for which excluded-middle fails. For, in order to resolve the normative paradoxes developed in §3.1, we required that Alpha be such that excluded-middle fails for the claim that it believes that $\ulcorner\beta\urcorner$ is true. However, if rejection were always the correct attitude to have in response to such propositions then we would be able to resurrect our normative paradox. Commitment, then, to CONSISTENCY, EVIDENCE and POSSIBILITY precludes us from saying that rejection is the correct cognitive response to all cases in which excluded-middle fails.

Given this fact, then, the proponent of REJECTION needs to provide a story about when it's okay to reject such a proposition and when some other attitude is appropriate. Here are some options. Although the proponent of REJECTION wants to say that rejection is always the correct cognitive response to cases of indeterminacy, perhaps she will say that rejection is not always the correct response to cases of indeterminacyⁿ, for some higher-order indeterminacy-type status. This option bifurcates. Perhaps the proponent of REJECTION will want to say that rejection is *never* the correct response to cases of indeterminacyⁿ, or perhaps it is only in certain cases like $\ulcorner\beta\urcorner$ that rejection is not the correct response.

Obviously this catalogue of options is rather schematic at this point. However, I think it's hard to see how the details could be filled in in a way which wouldn't involve rather *ad hoc* stipulations. *Prima facie* it seems quite implausible to hold that the correct response to failures of excluded-middle should be rejection in all cases *except* where this would lead us into normative paradox. Of course, this would get us out of trouble, but I think it's reasonable to ask for some independent reason that we should treat cases like $\ulcorner\beta\urcorner$ differently from other cases in which excluded-middle fails, and as I've already noted it's far from obvious what reason could be given here. This sort of worry doesn't apply to views that hold that rejection is never the correct response to certain types of higher-order indeterminacy. However, such views still raise the rather difficult question about why it is that rejection is the correct response to indeterminacy but not to, say, indeterminacyⁿ? What sort of plausible general principle could explain the difference in the cognitive significance of these statuses? This seems to me to be a rather tricky question for a proponent of REJECTION.

These considerations are clearly far from decisive. What they amount to is a challenge to the proponent of REJECTION to provide some plausible principled story about how we should respond to various cases in which excluded-middle fails that is compatible with resolving the normative paradoxes. Perhaps this can be done, but I think it's reasonable to be skeptical that any such account is available.

In contrast, a proponent of INDETERMINACY can provide what seems to me to be a simple and principled general story about how one should respond to different cases in which excluded-middle fails. The general principle underlying INDETERMINACY is, I suggest, the following: The correct cognitive response to any indeterminacy-type status is for there to be a mirroring indeterminacy-type status in one's doxastic state. Given this, we should accept the following schematic principle:

INDETERMINACYⁿ For any proposition ϕ , it is a consequence of the claim that one ought to believe

¹¹Field, for example, has endorsed this type of general rejectionism.

that ϕ is indeterminateⁿ, that one ought to be such that it is indeterminateⁿ whether one believes ϕ .

While it's true that we shouldn't respond to all cases in which excluded-middle fails in the same way, the proponent of INDETERMINACY, unlike the proponent of REJECTION, has what seems to me to be a principled story to tell about what attitudes one should have in different cases that is compatible with resolving the normative paradoxes. This seems to me to provide good *prima facie* reason to prefer INDETERMINACY to REJECTION.

Even, then, if we don't help ourselves to the claim that it's rational for it to be indeterminate whether Alpha believes that $\ulcorner\beta\urcorner$ is true, our earlier reflections on the normative paradoxes can be seen as providing support for giving up REJECTION in favor of INDETERMINACY.

Rejecting the Arguments for REJECTION

I've argued that we have good reason to prefer INDETERMINACY to REJECTION. REJECTION, however, is not without its own positive motivations. Here I'll take up what are, I think, the three clearest arguments in favor of REJECTION and show how they can be resisted.

Argument 1

Here is an argument which it might be thought strongly supports REJECTION.¹²

- (P1) One should reject any contradiction.
- (P2) The liar sentence entails a contradiction.
- (P3) The negation of the liar sentence entails a contradiction.
- (P4) If ϕ entails ψ then one should have at least as much confidence in ψ as in ϕ .
- (C) Therefore, one should reject both the liar sentence and its negation.

This argument, of course, extends to any sentence which, like the liar sentence, is such that both it and its negation entail a contradiction. Of course, we may want to extend the notion of indeterminacy to sentences that don't have this property, but in such cases considerations of uniformity could presumably be invoked.

The premiss that I reject is (P1). Certain contradictions are, according to a paracomplete theorist, indeterminate, for example, $T\ulcorner\lambda\urcorner \wedge \neg T\ulcorner\lambda\urcorner$. In these cases, I hold that it should be indeterminate whether one rejects the contradiction in question.

The question, then, is whether rejecting (P1) involves an unacceptable intuitive cost. Certainly (P1) is intuitive. It would, I think, be a significant drawback to the account I'm offering if there was nothing that I could say that could do justice to the intuitive pull of (P1). Ideally what we want

¹²See (Field 2003, 467) for this argument.

is (a) an alternative principle that can capture at least some of the intuitive force of (P1) and (b) an error-theory that can account for our mistakenly taking (P1) to be correct. I think that both of these desiderata can be met.

While we cannot hold that one should always reject a contradiction, we can hold that one should never determinately fail to reject a contradiction. Using this latter fact we can provide a plausible error-theory to account for our finding the former claim plausible. For, as noted earlier, the distinction between something being the case and it *determinately* being the case is not one to which our intuitions are sensitive, at least in advance of significant theoretical work. But if one ignores the distinction between something being the case and it determinately being the case, then the claim that one should never determinately fail to reject a contradiction will collapse to the claim that one should always reject a contradiction.¹³ And so one who thinks that one should never determinately fail to reject a contradiction should not be surprised if this correct principle was commonly confused for the incorrect principle that one should always reject a contradiction.

I don't deny that many will find the above argument in favor of REJECTION convincing, at least at first sight. But I don't think that the costs of rejecting premiss (P1) are all that significant. We can capture much of the intuitive force of this premiss. And we can explain why one who was not attuned to the possibility of doxastic indeterminacy would, on this basis, find (P1) plausible. This seems to me to take much of the sting out of rejecting (P1).

Argument 2

Here's another argument in favor of REJECTION. This argument appeals to degrees of belief. Assume that to believe a proposition ϕ is to have a degree of belief above a certain threshold τ , and to reject a proposition is to have a degree of belief below the co-threshold $1 - \tau$. Now we can argue as follows:

- (P1) ϕ entails $D(\phi)$.
- (P2) $D(\phi)$ entails ϕ .
- (P3) If ϕ entails ψ , then one should have at least as much confidence in ψ as in ϕ .
- (P4) One's degree of belief in ψ should be less than or equal to $1 -$ one's degree of belief in $\neg\psi$.

From (P1)-(P3) it follows that:

- (C1) One should have the same degree of belief in ϕ as in $D(\phi)$.

Given (P4) it follows that:

¹³To see that this is the case note that if one ignores the distinction between the conditions for something to be the case and the conditions for it to *determinately* be the case, then ϕ and $D\phi$ will be equivalent. Given this $\neg D\neg R\phi$ will be equivalent to $\neg\neg R\phi$ and so to $R\phi$.

(C2) If one has degree of belief over the threshold for $I\phi$, that is for $\neg D\phi \wedge \neg D\neg\phi$, then one should have a degree of belief below the co-threshold for $D\phi$ and for $D\neg\phi$.

and so given (C1):

(C3) If one has degree of belief over the threshold for $I\phi$, then one should have a degree of belief below the co-threshold for ϕ and for $\neg\phi$.

To adequately assess this argument we need to be a bit clearer about the notion of validity that a paracomplete theory endorses. In chapter 2 we characterized paracomplete validity as the preservation of value 1 in VF^+ models. This is one perfectly good notion of entailment that is available to a paracomplete theorist. Call this *weak entailment*. This is not, however, the only notion of entailment available. Another notion—call it *strong entailment*—is the following: ϕ strongly entails ψ just in case in every model ψ has a semantic value at least as great as the semantic value of ϕ .¹⁴ As it turns out the claim that ϕ strongly entails ψ is equivalent to the claim that the conditional $\phi \rightarrow \psi$ is weakly valid, that is to say that it has semantic value 1 in every model.

Given the distinction between strong and weak entailment there are two ways of understanding the above argument. We could either understand (P1)-(P3) as involving weak entailment:

(P1_w) ϕ weakly entails $D(\phi)$.

(P2_w) $D(\phi)$ weakly entails ϕ .

(P3_w) If ϕ weakly entails ψ then one should have at least as much confidence in ψ as in ϕ .

Or we could understand these premisses as involving strong entailment:

(P1_s) ϕ strongly entails $D(\phi)$.

(P2_s) $D(\phi)$ strongly entails ϕ .

(P3_s) If ϕ strongly entails ψ then one should have at least as much confidence in ψ as in ϕ .

Either way the argument can be resisted. If the argument is understood in terms of weak entailment then (P1_w) and (P2_w) both hold. I claim, however, that we should reject (P3_w). Instead, we should only accept (P3_s). It is strong entailment, not weak entailment, that should be thought of as providing a normative constraint on our degrees of belief. However, accepting (P3_s) does not provide adequate materials for the argument in favor of REJECTION. For, while (P1_w) holds, (P1_s) does not. ϕ does not, in general, strongly entail $D\phi$. In fact, the only cases in which it does are ones in which ϕ is valid. So, the only case in which one is required by logic to have the same degree of belief in ϕ and $D\phi$ is when ϕ is a logical validity. Such cases don't provide any trouble, since presumably one shouldn't believe that such sentences are indeterminate.

¹⁴More generally, we can say that a set of sentences Γ strongly entails ψ iff in every model the semantic value of ψ is at least as great as the greatest lower bound of the semantic values of members of Γ .

It is certainly true that we want a notion of entailment that constrains our degrees of belief in the manner specified in (P3). However, adopting the model theory developed by Field for the treatment of indeterminacy does not force us to accept REJECTION in order to meet this desideratum. We can hold that the relevant notion is strong entailment. A minimal point, then, is that a defender of INDETERMINACY does have the resources available to resist this argument in favor of REJECTION

But such a defender can say something stronger. For I think that there is good independent reason to think that it is strong entailment that should be thought of as having a normative role to play in constraining the degrees of belief of rational agents.

It is often said that belief aims at truth or has as its goal truth. A natural corollary to this thought is the following. The reason why our beliefs should be constrained in the manner described by (P3) is that truth is preserved under entailment. For, if our goal as believers is to believe the truth, then, given that whenever ϕ is true ψ is true, whatever confidence we have in ϕ should also be invested in ψ .

The important point is that if this is the justification for (P3) then it is strong entailment and not weak entailment that is the relevant notion. For while we can say that if ϕ strongly entails ψ , then if ϕ is true then ψ is true (and of logical necessity), we cannot say the same thing of weak entailment. Let me explain. As noted above, the claim that ϕ strongly entails ψ is equivalent to the claim that $\phi \rightarrow \psi$ is valid. The latter is equivalent to the claim that $T^\Gamma \phi^\neg \rightarrow T^\Gamma \psi^\neg$ is valid. So, given that ϕ strongly entails ψ , it follows of logical necessity that if ϕ is true then ψ is true. We cannot, however, say the same for weak validity. It does not follow from the fact that an inference preserves semantic value 1 that it preserves truth; that is we cannot infer from $\phi \models \psi$ to $\models T^\Gamma \phi^\neg \rightarrow T^\Gamma \psi^\neg$. Since ϕ and $T^\Gamma \phi^\neg$ are intersubstitutable, the explanation for this is that the deduction theorem fails for weak-validity. And the deduction theorem must fail, for otherwise the Curry paradox could not be given an adequate solution.¹⁵

It seems to me, then, that the most natural justification for (P3) motivates understanding this principle in terms of strong validity. Not only can the argument under consideration be resisted, but such resistance is independently motivated.

Argument 3

Here's a final argument in favor of REJECTION.

The hope of providing an informative analysis of indeterminacy, at least as it applies to the liar paradox, is slim. Certainly standard analyses that have been thought promising in the case of vagueness are hopeless when we are dealing with semantic paradoxes. How, then, one might ask, are we to understand what it is for a proposition to be indeterminate?

A proponent of REJECTION can say the following. While we cannot provide an analysis of indeterminacy, we can come to understand the concept by seeing the role that it plays in our cognitive lives. Indeterminacy would in this way be like objective chance.¹⁶

¹⁵See (Field 2008, ch. 19) for a discussion of the relationship between the deduction theorem for weak validity and the Curry paradox.

¹⁶It is plausible to think, first, that objective chance cannot be analyzed in more basic terms, and, second, that at

I, however, can say no such thing. For I think that we need to *use* the concept of indeterminacy in order to characterize the distinctive cognitive role of indeterminacy.

REJECTION, then, gives us an independent grip on indeterminacy that INDETERMINACY does not. And this, so the argument goes, is a significant advantage of REJECTION over INDETERMINACY.

It must be admitted that it is a cost of my view that it deprives us of this independent grip on the concept of indeterminacy. Nonetheless, I suggest that on careful inspection the asymmetry between myself and the proponent of REJECTION is not that great.

To see this point, first recall that an adequate treatment of the liar paradox that avails itself of the notion of indeterminacy requires that there be a hierarchy of non-equivalent indeterminacy-type operators. The question arises, then, for each $n > 1$, what attitude should one have towards a proposition ϕ that one ought to take to be indeterminateⁿ? Let's focus on a simple case. Let ϕ be a proposition that one ought to believe is indeterminately indeterminate. What attitude should one have towards ϕ ?

Here are the reasonable responses that I think are available to a proponent of REJECTION:

- (i) The proponent of REJECTION may say that in every such case one should reject ϕ and its negation.
- (ii) The proponent of REJECTION may say that in at least some such cases one should be such that it is indeterminately indeterminate whether one believes ϕ

(i) can be motivated as follows. A proponent of REJECTION holds that it is a consequence of the claim that one ought to believe $\neg D\phi$ and $\neg D\neg\phi$ that one ought to reject ϕ and its negation. One who accepts this, should, I think, be inclined to also accept the more general claim:

REJECTION* For any proposition ϕ , it is a consequence of the claim that one ought to reject $D\phi$ and $D\neg\phi$ that one ought to reject ϕ and its negation.¹⁷

Using REJECTION* we can provide an argument that $OBII\phi \models OR\phi \wedge OR\neg\phi$.

To argue for this it will suffice to argue for the following claims:

$$(25) \quad OBII\phi \models OR\neg I\phi$$

$$(26) \quad OR\neg I\phi \models OR\phi \wedge OR\neg\phi$$

least a large part of our understanding of objective chance consists in our knowing that whatever objective chance is it should play something like the following role in our cognitive lives: If one is rational and one believes that the chance of ϕ occurring is x and one has no additional information about ϕ then one will have credence x in ϕ . See Lewis (1986) for an argument that this provides the foundation for our understanding of objective chance. See (Field 2003, 479) for a comparison between indeterminacy and chance in this respect.

¹⁷Field, for example, would accept this more general statement. For, as noted, he thinks that it is a rational requirement that one's degree of belief in ϕ be the same as one's degree of belief in $D\phi$.

(25) is an obvious consequence of REJECTION.

(26) can be established by the following simple argument: First note that $\neg I\phi$ is equivalent to $D\phi \vee D\neg\phi$. We have then $OR\neg I\phi \models OR(D\phi \vee D\neg\phi)$. The following strikes me as non-negotiable norm governing rejection: $OR(\gamma \vee \psi) \models OR\gamma \wedge OR\psi$. In particular then we have $OR(D\phi \vee D\neg\phi) \models ORD\phi \wedge ORD\neg\phi$. So by transitivity of entailment we have $OR\neg I\phi \models ORD\phi \wedge ORD\neg\phi$. By REJECTION* we have $ORD\phi \wedge ORD\neg\phi \models OR\phi \wedge OR\neg\phi$. And so, finally, we have $OR\neg I\phi \models OR\phi \wedge OR\neg\phi$, which is (26).

The only reason that I can see for a proponent of REJECTION to not accept (i) (and so to try to find a way of resisting this argument) would be if she wanted to treat the proposition expressed by $\ulcorner\beta\urcorner$, or some other similar proposition, as being indeterminately indeterminate, as a way addressing the normative paradox developed in §3.1. In this case, however, in order to avoid resurrecting the normative paradox, the proponent of REJECTION should hold that the correct response to the perceived higher-order indeterminacy is for it to be indeterminately indeterminate whether the agent believes the proposition in question. Thus if the proponent of REJECTION doesn't accept (i), she should accept (ii).

Now in either case we can argue that the explanatory asymmetry between REJECTION and INDETERMINACY is not as great as it might at first appear.

If, on the one hand, the proponent of REJECTION decides to accept (i), then we can argue as follows. It is true that REJECTION gives us a grip on indeterminacy in terms which do not presuppose an understanding of this concept. Nonetheless REJECTION does not distinguish between indeterminacy and indeterminate indeterminacy, for there are equivalent principles governing this latter operator. While being indeterminate and being indeterminately indeterminate are not the same we cannot understand the difference between them solely by appeal to the attitudes of rational agents towards propositions not involving indeterminacy.

If, on the other hand, the proponent of REJECTION decides to accept (ii), then we can argue as follows. While it is true that REJECTION does provide us with an independent grip on indeterminacy, the proponent of REJECTION must still allow that there are other indeterminacy-type statuses, in particular indeterminate indeterminacy, whose cognitive role cannot be specified in terms that don't presuppose an understanding of that very notion of higher-order indeterminacy.

In either case, then, the conclusion to be drawn is that the proponent of REJECTION can't provide us with an understanding of indeterminate indeterminacy, and how it differs from indeterminacy, by appeal solely to an independently specifiable cognitive role. In the first case, we can understand the cognitive role without having an independent grip on indeterminate indeterminacy. However, the cognitive role is insufficient to distinguish indeterminate indeterminacy from indeterminacy. In the second case, the cognitive role does distinguish the higher-order status from first-order indeterminacy. However, we can't understand this cognitive role without already having an understanding of indeterminate indeterminacy.

How, then, can a proponent of REJECTION account for our understanding of indeterminate indeterminacy, and the difference between this status and simple indeterminacy? This is difficult question. In outline, what I think the proponent of REJECTION will have to say is that our understanding of indeterminate indeterminacy and the way in which it differs from indeterminacy comes from our grasping (a) the different paradigm cases in which the higher-order operator ap-

plies and (b) the difference in its logical behavior from other indeterminacy-type operators. The difference between the proponent of REJECTION and myself is that I don't have, in addition to these resources, facts about cognitive significance to help explain our understanding of various indeterminacy-type operators. Now, it is certainly preferable *ceteris paribus* to have available more resources in order to explain the primitives of one's theory. But giving up one's right to appeal to facts about attitudes in one's explanation of indeterminacy does not strike me as intolerable given that one must, in any case, avail oneself of other resources in order to fully account for our understanding of related concepts. It is a cost of accepting INDETERMINACY that one can no longer appeal to attitudes in order to explain indeterminacy. But if, as I've argued, there are arguments that point strongly in favor of INDETERMINACY, then this is a cost, I think, worth incurring.

6.1.2 Paracomplete Cognitive Significance: Quantitative Attitudes

A Negative Argument

The qualitative principle REJECTION has a quantitative counterpart QUANTITATIVE REJECTION:

QUANTITATIVE REJECTION For any proposition ϕ , it is a consequence of the claim that one ought to be such that $Cr(I\phi) = x$ that one ought to be such that $Cr(\phi) + Cr(\neg\phi) \leq 1 - x$.¹⁸

Given certain not implausible auxiliary assumptions, we can see REJECTION as an instance of this more general principle. In particular, let's us assume that to believe a proposition is to have a credence in the proposition at or above some threshold value τ such that $\tau > 0.5$. We can then say that to reject a proposition is to have credence at or below the co-threshold $\rho = 1 - \tau$. Given this connection between credence and belief/rejection, it then clearly follows from QUANTITATIVE REJECTION that, given that an agent ought to believe that ϕ is indeterminate, the agent ought to reject both ϕ and its negation. For given our auxiliary principles, the assumption that the agent ought to believe that ϕ is indeterminate is just the assumption that the agent ought to have credence at or above τ in the claim $I\phi$. And given QUANTITATIVE REJECTION, we then have that the agent ought to be such that the combined value of its credences in ϕ and in $\neg\phi$ is at or below $1 - \tau$. But this clearly entails that the agent ought to have credence at or below $1 - \tau$ in ϕ and credence at or below $1 - \tau$ in $\neg\phi$. And, given our auxiliary assumptions, this is tantamount to the claim that the agent ought to reject both ϕ and $\neg\phi$.

QUANTITATIVE REJECTION can be motivated by appeal to other *prima facie* plausible principles:

CREDESCENCE₁ $\models \phi \rightarrow \psi \Rightarrow O(Cr(\phi) \leq Cr(\psi))$

CREDESCENCE₂^w $O(Cr(\neg\phi) \leq 1 - Cr(\phi))$

CREDESCENCE₃ $O(Cr(\phi \vee \psi) = Cr(\phi) + Cr(\psi) - Cr(\phi \wedge \psi))$

¹⁸Read: $O(Cr(I\phi) = x) \models O(Cr(\phi) + Cr(\neg\phi) \leq 1 - x)$. See Field (2003) and Field (2008).

CREDESCENCE₄ $O(Cr(\phi) = Cr(D\phi))$

To see how these can be used to justify QUANTITATIVE REJECTION, first note that $I\phi$ is equivalent to (indeed, is a mere abbreviation for) $\neg D\phi \wedge \neg D\neg\phi$. Assume that the agent ought to be such that $Cr(\neg D\phi \wedge \neg D\neg\phi) = x$. By CREDENCE₁, we get that if the agent is rational then: $Cr(\neg(D\phi \vee D\neg\phi)) = x$. By CREDENCE₁^w, we get that if the agent is rational then: $Cr(D\phi \vee D\neg\phi) \leq 1 - x$. Assuming, as one might think is reasonable, that if one is rational then: $Cr(D\phi \wedge D\neg\phi) = 0$, then CREDENCE₃ gives us that if the agent is rational then: $Cr(D\phi) + Cr(D\neg\phi) \leq 1 - x$. And so finally, from CREDENCE₄ we get that if the agent is rational then: $Cr(\phi) + Cr(\neg\phi) \leq 1 - x$.¹⁹

In the previous section, we saw how our solution to the paradox of belief could be used to motivate rejecting REJECTION. In this section, I'll show how the solution to the paradox of credence provides materials for a similar argument against QUANTITATIVE REJECTION.

In §3.2, we showed that, given classical assumptions, CREDENCE₁ and CREDENCE₂ are classically inconsistent with POSSIBILITY:

In §4.2.1, we showed how this paradox can be resolved if we allow that excluded-middle may fail for certain claims about an agent's credences.

Now, it's worth noting that the paradox as stated won't have much force for a proponent of QUANTITATIVE REJECTION. For, insofar as the proponent of QUANTITATIVE REJECTION thinks that there are cases in which an agent may rationally have positive credence in $I\phi$, she will want to allow that in such a case an agent may rationally fail to satisfy CREDENCE₂. The reason for this is that in order to meet the requirements imposed by QUANTITATIVE REJECTION, the agent's credences in ϕ and $\neg\phi$ must sum up to some value less than 1, and this can only hold if the agent does not satisfy CREDENCE₂.

Nonetheless, as noted in §3.2, the argument that generates the incompatibility between CREDENCE₁, CREDENCE₂ and POSSIBILITY, also shows that that CREDENCE₁ and POSSIBILITY are incompatible with the weaker principle:

CREDESCENCE₂^w $O(Cr(\neg\phi) \leq 1 - Cr(\phi))$

And this latter principle is one that a proponent of QUANTITATIVE REJECTION *should* accept. We can use this alternative normative paradox, and the solution to this paradox provided in §4.2.1, to provide an argument against QUANTITATIVE REJECTION.

The argument takes a very similar form to the argument provided in the previous section against REJECTION. In §4.2.1, we showed that an agent could satisfy the requirements imposed by CREDENCE₁ and CREDENCE₁^w, given that excluded-middle fails for the claim: $Cr(T(\ulcorner\beta'\urcorner)) > 0.9$. Since any paracomplete theorist should accept CREDENCE₁ and CREDENCE₂^w, such theorists should hold that if Alpha is rational it will be such that excluded-middle fails for the claim that $Cr(T(\ulcorner\beta'\urcorner)) > 0.9$. I take it—for the same reasons articulated in the case of qualitative belief in §6.1.1—that a rational way for Alpha to meet this requirement is for it to be indeterminate whether it's credence in the truth of $\ulcorner\beta'\urcorner$ is over 0.9. (As in the case of qualitative belief, this is, indeed, the

¹⁹See Field (2003).

status assigned in the VF^+ models in which we've represented Alpha's credal state.) I'll further assume that Alpha will be certain of $I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)$ just in case this proposition obtains. (As in the case of qualitative belief, this assumption is satisfied in the VF^+ models in which we've represented Alpha's credal state.) That is, I assume:

$$(27) \quad Cr(I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)) = 1 \leftrightarrow I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)$$

From (27), and our assumption that Alpha is such that $I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)$, we have: $Cr(I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)) = 1$. Since Alpha's being such that $I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)$ is a rational way for Alpha to meet the requirements imposed by CREDENCE₁ and CREDENCE₂^w, and since, I take it, Alpha is not irrational in virtue of having the introspective powers ascribed in (27), it follows that it is rational for Alpha to be such that $Cr(I(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)) = 1$. Assuming, next, that Alpha meets the requirements imposed by QUANTITATIVE REJECTION, we have: $Cr(\neg(Cr(T(\ulcorner\beta'\urcorner)) > 0.9)) = 0$. But this lands us right back in normative paradox. To see this we simply need to rehearse Case 2 from §3.2. There it was established that Alpha is guaranteed to be in violation of CREDENCE₁, on the assumption that its credence in the truth of $\ulcorner\beta'\urcorner$ is less than or equal to 0.9.

What this tells us is that if, in addition to CREDENCE₁ and CREDENCE₂^w, we accept QUANTITATIVE REJECTION, we will once again have to admit that there are situations in which an antecedently rational agent will be doomed to irrationality. But since this is exactly what POSSIBILITY denies, this is sufficient to show that CREDENCE₁, CREDENCE₂^w and QUANTITATIVE REJECTION are incompatible with POSSIBILITY. As in the earlier case, the derivation of a contradiction from the assumption that Alpha meets the requirements imposed by CREDENCE₁, CREDENCE₂^w and QUANTITATIVE REJECTION doesn't appeal to any inference principles that are suspect by paracomplete lights. Unlike in the case of $\ulcorner\beta'\urcorner$, then, paracomplete resources are of no use in blocking the contradiction.

A paracomplete theorist must, then, choose between CREDENCE₁, CREDENCE₂^w, QUANTITATIVE REJECTION and POSSIBILITY. And, as in our earlier case, it seems to me obvious which of these should go. CREDENCE₁, CREDENCE₂^w and POSSIBILITY are much more central to our understanding of credal states than QUANTITATIVE REJECTION. While QUANTITATIVE REJECTION may have some *prima facie* plausibility, its incompatibility with these fundamental principles governing credence show that it is ultimately unacceptable.

As noted earlier, there are plausible principles that are jointly sufficient for QUANTITATIVE REJECTION. If, then, we're going to reject QUANTITATIVE REJECTION, we need to reject at least one of these principles. I've already argued, however, that a paracomplete theorist should reject CREDENCE₄. We shouldn't hold that rationality requires an agent to have the same credence both in a claim ϕ and in $D\phi$.

Finally, it's worth noting that it is a *prima facie* unattractive feature of QUANTITATIVE REJECTION that it requires rejection of CREDENCE₂. It is natural to think that it is *always* appropriate to express the attitude of rejection towards a proposition ϕ with the endorsement of $\neg\phi$. Indeed, it is natural to think that this is a characteristic feature of negation. The proponent of QUANTITATIVE REJECTION needs to deny this. It is a noteworthy attractive feature of rejecting QUANTITATIVE REJECTION, then, that it frees us up to once again endorse CREDENCE₂.

A Positive Proposal

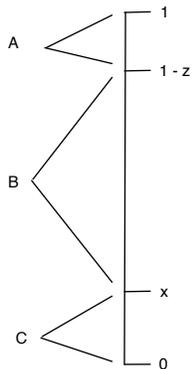
QUANTITATIVE REJECTION provides us with an account of how credence in $I\phi$ may rationally constrain one's credence towards ϕ . Unfortunately, it provides us with the wrong account of this relation. What rational constraints does credence in $I\phi$ impose on one's credence in ϕ ?

Here, things are actually easier if we formulate our answer in terms of the constraints that certain credences in $D\phi$ and $D\neg\phi$ impose on one's credences in ϕ . As a counterpart to INDETERMINACY in the case of qualitative belief, I propose the following principle for quantitative beliefs:

QUANTITATIVE INDETERMINACY For any proposition ϕ , it is a consequence of the claim that one ought to be such that $Cr(D\phi) = x$ and $Cr(D\neg\phi) = z$, that one ought to be such that the following all hold:

- (i) $\forall y(y < x \rightarrow D(Cr(\phi) \neq y))$
- (ii) $\forall y(y > 1 - z \rightarrow D(Cr(\phi) \neq y))$
- (iii) $x \neq 1 - z \rightarrow (ICr(\phi) = x \wedge ICr(\phi) = 1 - z)$
- (iv) $x = 1 - z \rightarrow (DCr(\phi) = x)$

Let's spell out a bit what this says. Consider the following representation of an agent's possible credences ranging from 0 to 1. We'll assume that the agent has credence x in $D\phi$ and that the agent has credence z in $D\neg\phi$.



QUANTITATIVE INDETERMINACY tells us that, for a rational agent, x provides an upper-bound, so that all possible credences lower than x , i.e., all of the credences in the interval C, are determinately not the credence the agent has in ϕ . It also tells us that $1 - z$ provides a lower-bound, so that all possible credence above $1 - z$, i.e., all credences in the interval A, are determinately not the credence the agent has in ϕ . Finally, QUANTITATIVE INDETERMINACY tells us that that if the interval B is point sized, then that point, i.e., x , is determinately the agent's credence in ϕ , while

if B is not point sized, then for each of the end points of B , it is indeterminate whether it is the agent's credence in ϕ .²⁰

As in the case of INDETERMINACY, we can use our VF^+ models to show that QUANTITATIVE INDETERMINACY, unlike QUANTITATIVE REJECTION, will not reinstate our normative paradox. Let \mathcal{M} be the class of VF^+ models of the type outlined in §4.2.1, with the following property: for any set of points Q such that $\mathcal{P}(Q) > 0$, there is some point $\delta \in Q$, such that $\mathcal{P}(\{\delta\}) > 0$. We can show that the following all hold:

$$(28) \quad Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \models_{\mathcal{M}} \forall y(y < x \rightarrow D(Cr(\phi) \neq y))$$

$$(29) \quad Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \models_{\mathcal{M}} \forall y(y > 1 - z \rightarrow D(Cr(\phi) \neq y))$$

$$(30) \quad Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \models_{\mathcal{M}} x \neq 1 - z \rightarrow (ICr(\phi) = x \wedge ICr(\phi) = 1 - z)$$

$$(31) \quad Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \models_{\mathcal{M}} x = 1 - z \rightarrow (DCr(\phi) = x \wedge DCr(\neg\phi) = z)$$

Proof of (28)

Assume: $\llbracket Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \rrbracket^g = 1$. We want to show: $\llbracket \forall y(y < x \rightarrow D(Cr(\phi) \neq y)) \rrbracket^g = 1$. To do this, let g' be a y -variant of g , and assume that $g'(y) < g'(x)$. We want to show that for every stage α in our VF^+ model, $\llbracket D(Cr(\phi) \neq y) \rrbracket_{\alpha}^{g'} = 1$. $D(Cr(\phi) \neq y)$ is equivalent to $Cr(\phi) \neq y \wedge \neg(Cr(\phi) \neq y \rightarrow Cr(\phi) = y)$. So we want to show (a) $\llbracket Cr(\phi) \neq y \rrbracket_{\alpha}^{g'} = 1$, and (b) $\llbracket \neg(Cr(\phi) \neq y \rightarrow Cr(\phi) = y) \rrbracket_{\alpha}^{g'} = 1$.

Note the following fact: (i) $\mathcal{P}(\Sigma^{\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket_{\alpha}=1}) = g'(x)$. But from (i), (a) clearly follows. For given $\mathcal{P}(\Sigma^{\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket_{\alpha}=1}) = g'(x)$, it follows that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) \geq g'(x)$. And since $g'(y) < g'(x)$, it follows that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) > g'(y)$. And this means that for every $Q \subset \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1} \cup Q) \neq g'(y)$. And this suffices to establish (a).

To see that (b) holds, note that if α is a successor ordinal, then (b) follows by applying the above argument for (a) to the preceding ordinal. And if α is a limit ordinal, (b) follows given that it holds for all the preceding successor ordinals.

Proof of (29)

Assume: $\llbracket Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \rrbracket^g = 1$. We want to show: $\llbracket \forall y(y > 1 - z \rightarrow D(Cr(\phi) \neq y)) \rrbracket^g = 1$. To do this, let g' be a y -variant of g , and assume that $g'(y) > 1 - g'(z)$. We want to show that for every stage α in our VF^+ model, $\llbracket D(Cr(\phi) \neq y) \rrbracket_{\alpha}^{g'} = 1$. $D(Cr(\phi) \neq y)$ is equivalent to $Cr(\phi) \neq y \wedge \neg(Cr(\phi) \neq y \rightarrow Cr(\phi) = y)$. So we want to show (a) $\llbracket Cr(\phi) \neq y \rrbracket_{\alpha}^{g'} = 1$, and (b) $\llbracket \neg(Cr(\phi) \neq y \rightarrow Cr(\phi) = y) \rrbracket_{\alpha}^{g'} = 1$.

²⁰Why not say, for each point in B , that it is indeterminate whether that is the agent's credence? The reason is that in many cases this won't be true for VF^+ agents. For, even if $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) = x$ and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=0}) = z$, it doesn't follow that for any point b in B there is some subset Q of $\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$, such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1} \cup Q) = b$. For example, the points over which the agent's credences are divided may be such that dividing up $\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ will either overshoot or undershoot b .

Note the following fact: (i) $\mathcal{P}(\Sigma^{\llbracket \neg\phi \wedge \neg(\neg\phi \rightarrow \phi) \rrbracket_{\alpha}=1}) = g'(z)$. But from (i), (a) clearly follows. We know that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) + \mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha}=1}) = 1$. And since, by (i), we know $\mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha}=1}) \geq g'(z)$, It follows that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) = 1 - \mathcal{P}(\Sigma^{\llbracket \neg\phi \rrbracket_{\alpha}=1}) \leq 1 - g'(z)$. And so, since $g'(y) > 1 - g'(z)$, we have $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) < g'(y)$. This assures us that (a) holds.

To see that (b) holds, we can once again apply the argument for (a) to the preceding ordinal if α is a successor. And, if α is a limit ordinal the fact that (b) holds for all preceding successor ordinals assures us that (b) will hold for α as well.

Proof of (30)

Assume: $\llbracket Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \rrbracket^g = 1$. We want to show: $\llbracket x \neq 1 - z \rightarrow (ICr(\phi) = x \wedge ICr(\phi) = 1 - z) \rrbracket^g = 1$. Assume: $g(x) \neq 1 - g(z)$. We want to show: (a) $\llbracket ICr(\phi) = x \rrbracket^g = 1$, and (b) $\llbracket ICr(\phi) = 1 - z \rrbracket^g = 1$. To show that (a) and (b) hold we want to show: (a') $\llbracket Cr(\phi) = x \rrbracket^g = 1/2$, and (b') $\llbracket Cr(\phi) = 1 - z \rrbracket^g = 1/2$.

To show that (a') and (b') hold, first note that there will be some ordinal γ , such that $\llbracket D\phi \rrbracket_{\gamma}^g = 1$ just in case $D\phi$ has semantic value 1 at every ordinal in our VF^+ model, and $\llbracket D\neg\phi \rrbracket_{\gamma}^g = 1$ just in case $D\neg\phi$ has semantic value 1 at every ordinal in our VF^+ model.²¹ Let's label the set of points in our model at which $D\phi$ has value 1 at every ordinal: X , and the set of points at which $D\neg\phi$ has value 1 at every ordinal: Y . By our assumption we have: $\llbracket Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \rrbracket^g = 1$. This tells us that $\mathcal{P}(\Sigma^{\llbracket D\phi \rrbracket_{\gamma}=1}) = g(x)$ and $\mathcal{P}(\Sigma^{\llbracket D\neg\phi \rrbracket_{\gamma}=1}) = g(z)$. Since $X = \Sigma^{\llbracket D\phi \rrbracket_{\gamma}=1}$, and $Y = \Sigma^{\llbracket D\neg\phi \rrbracket_{\gamma}=1}$, we have $\mathcal{P}(X) = g(x)$ and $\mathcal{P}(Y) = g(z)$. For any ordinal α , we know that for every $\delta \in X$, $\llbracket \phi \rrbracket_{\alpha}^g = 1$, and for every $\delta \in Y$, $\llbracket \phi \rrbracket_{\alpha}^g = 0$. So this assures us that for any α , $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) \geq g(x)$, and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=0}) \geq g(z)$. What I'll now argue is that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) = g(x)$, and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=0}) = g(z)$. Since $g(x) \neq 1 - g(z)$, this assures us that for every α there is some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ such that $\mathcal{P}(\Sigma_{\alpha}^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) = g(x)$ (viz., \emptyset) and some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ such that $\mathcal{P}(\Sigma_{\alpha}^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) \neq g(x)$ (viz., $\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$). This, then, suffices to establish (a'). We can also be assured that there is some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ such that $\mathcal{P}(\Sigma_{\alpha}^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) \neq 1 - g(z)$ (viz., \emptyset) and some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$ such that $\mathcal{P}(\Sigma_{\alpha}^{\llbracket \phi \rrbracket_{\alpha}=1} \cup \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}) = 1 - g(z)$ (viz., $\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$). This, then, suffices to establish (b').

To show that for any α , $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=1}) = g(x)$ and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha}=0}) = g(z)$, it will suffice to show that there is no $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$, such that $\mathcal{P}(Q) > 0$, and such that either ϕ has semantic value 1 for all $\delta \in Q$, or ϕ has semantic value 0 for all $\delta \in Q$. Assume, for reductio, that there is some, $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha}=1/2}$, such that $\mathcal{P}(Q) > 0$, and such that ϕ has semantic value 1 for all $\delta \in Q$. (It will be clear how the corresponding case would for for the assumption that ϕ has semantic value 0 at these points.) By assumption, we are dealing with a model such that any set with positive measure has a point with positive measure. So we know that there is some δ such that $\llbracket \phi \rrbracket_{\alpha}^{\delta} = 1$ and such that $\mathcal{P}(\{\delta\}) > 0$.

²¹This is a consequence of the, so-called, FUNDAMENTAL THEOREM. See Field (2008).

We can argue that at $\alpha + 1$, $\llbracket \phi \rrbracket_{\alpha+1}^\delta = 0$. For if $\llbracket \phi \rrbracket_{\alpha+1}^\delta \neq 0$, then $\llbracket D\phi \rrbracket_{\alpha+1}^\delta \neq 0$, i.e., $\llbracket \phi \wedge \neg(\phi \rightarrow \neg\phi) \rrbracket_{\alpha+1}^\delta \neq 0$. But this is incompatible with $\llbracket Cr(D\phi) = x \rrbracket_{\alpha+1}^g = 1$. For if $\llbracket D\phi \rrbracket_{\alpha+1}^\delta = 1$, then $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha+1}=1}) > g(x)$, while if $\llbracket D\phi \rrbracket_{\alpha+1}^\delta = 1$, then there would be some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_{\alpha+1}=1/2}$, such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha+1}=1} \cup Q) > g(x)$. But given that $\llbracket \phi \rrbracket_{\alpha+1}^\delta = 0$, we can now similarly argue that at $\alpha + 2$ we must have $\llbracket \phi \rrbracket_{\alpha+2}^\delta = 1$. For, otherwise, we couldn't have $\llbracket Cr(D\neg\phi) = z \rrbracket_{\alpha+2}^g = 1$. (I leave the details of this part of the argument for the reader, since they are only trivially different from the $\alpha + 1$ case.)

What we have seen is that on the assumption that there is some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_\alpha=1/2}$ such that $\mathcal{P}(Q) > 0$ and such that ϕ has semantic value 1 for all $\delta \in Q$, it follows that there is a point δ with positive measure such that at that point ϕ alternates between values 1 and 0 as our ordinals proceed. But given this, it follows that at the first limit ordinal, λ , following α , $\llbracket D\phi \rrbracket_\lambda^\delta = 1/2$. And so it follows that there is some $Q \subseteq \Sigma^{\llbracket \phi \rrbracket_\lambda=1/2}$, such that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_{\alpha+1}=1} \cup Q) > g(x)$. But this is incompatible with $\llbracket Cr(D\phi) = x \rrbracket = 1$.

This shows, then, that for every set Q with positive measure, if $Q \subseteq \Delta - X - Y$, then for every ordinal α and every $\delta \in Q$ $\llbracket \phi \rrbracket_\alpha^\delta \neq 1$. A similar argument will show that for every ordinal α and every $\delta \in Q$ $\llbracket \phi \rrbracket_\alpha^\delta \neq 0$. It follows that for every ordinal α , $\llbracket \phi \rrbracket_\alpha^\delta = 1$ just in case $\delta \in X$, and $\llbracket \phi \rrbracket_\alpha^\delta = 0$ just in case $\delta \in Y$. This, then, suffices to establish that for every α , $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha=1}) = g(x)$, and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha=0}) = g(z)$.

Proof of (31)

Assume: $\llbracket Cr(D\phi) = x \wedge Cr(D\neg\phi) = z \rrbracket^g = 1$. We want to show: $\llbracket x = 1 - z \rightarrow DCr(\phi) = x \rrbracket^g = 1$. Assume: $g(x) = 1 - g(z)$. We want to show: $\llbracket DCr(\phi) = x \rrbracket^g = 1$. To show this, we want to show: $\llbracket Cr(\phi) = x \rrbracket^g = 1$.

Let α be an arbitrary ordinal. We'll show: $\llbracket Cr(\phi) = x \rrbracket_\alpha^g = 1$. We know that $\llbracket Cr(D\phi) = x \rrbracket_\alpha^g = 1$ and $\llbracket Cr(D\neg\phi) = z \rrbracket_\alpha^g = 1$. This means that $\mathcal{P}(\Sigma^{\llbracket D\phi \rrbracket_\alpha=1}) = g(x)$ and $\mathcal{P}(\Sigma^{\llbracket D\neg\phi \rrbracket_\alpha=1}) = g(z)$. And since $g(x) + g(z) = 1$, it follows that $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha=1}) = g(x)$ and $\mathcal{P}(\Sigma^{\llbracket \phi \rrbracket_\alpha=1/2}) = 0$. And thus it follows that $\llbracket Cr(\phi) = x \rrbracket_\alpha^g = 1$.

(28) - (31) are sufficient to assure us that Alpha can, indeed, meet the requirements imposed by CREDENCE₁, CREDENCE₂ and QUANTITATIVE INDETERMINACY. Since in any VF^+ model all of the requirements imposed by CREDENCE₁ and CREDENCE₂ will be met, it follows from (28) - (31) that in any VF^+ model in which the measure over points is such that any set with positive measure has points with positive measure all of the requirements imposed by CREDENCE₁, CREDENCE₂ and QUANTITATIVE INDETERMINACY will be met.²²

An important point to highlight is that just as QUANTITATIVE REJECTION gives us REJECTION as a special case under the assumption that belief and rejection can be understood in terms of thresholds of credences, so too does QUANTITATIVE INDETERMINACY give us INDETERMINACY

²²To see how this works, simply apply, *mutatis mutandis*, the argument, outlined in §6.1, that showed that $Bl\phi \models_{\mathcal{M}} IB\phi$ is sufficient to assure us that Alpha can meet all the requirements imposed by CONSISTENCY, EVIDENCE and INDETERMINACY.

as a special case. At least it does, given other plausible rational constraints on credence that can be shown to hold in the class of VF^+ models in which we've represented paracomplete credences.

To see this, let's assume that an agent believes a proposition ϕ just in case their credence in ϕ is at or above some threshold $\tau > 0.5$ and that the agent rejects ϕ just in case their credence in ϕ is at or below the co-threshold $\rho = 1 - \tau$. Assume, further, that the agent ought to believe $I\phi$, i.e., assume that the agent ought to be such that $Cr(\neg D\phi \wedge \neg D\neg\phi) = x \geq \tau$. We'll show that, on the assumption that the agent meets this obligation and the obligations imposed by QUANTITATIVE INDETERMINACY and a few other reasonable rational principles, it will be indeterminate whether the agent believes ϕ , i.e., it will be indeterminate whether $Cr(\phi) \geq \tau$.

For concreteness, we'll assume that $\tau = 0.8$, but nothing essential turns on this. We assume that our agent ought to believe that ϕ is indeterminate. So, assuming that the agent meets this obligation, we have: $Cr(\neg D\phi \wedge \neg D\neg\phi) \geq 0.8$. Assuming that the agent satisfies CREDENCE₁, we have $Cr\neg(D\phi \vee D\neg\phi) \geq 0.8$. And assuming that the agent satisfies CREDENCE₂, we have: $1 - Cr(D\phi \vee D\neg\phi) \geq 0.8$. And so, we have: $Cr(D\phi \vee D\neg\phi) \leq 0.2$. Now, a plausible general constraint that will be satisfied by any VF^+ agent is the following: $O(Cr(D\phi \vee D\neg\phi) = x) \models O(Cr(D\phi) + Cr(D\neg\phi) = x)$.²³ Assuming that our agent satisfies this general constraint, we have: $Cr(D\phi) + Cr(D\neg\phi) \leq 0.2$. From this latter claim it follows that: (i) $Cr(D\phi) < 0.8$ and (ii) $1 - Cr(D\neg\phi) \geq 0.8$. From (i) and (ii), it follows that if the agent meets the requirements imposed by QUANTITATIVE INDETERMINACY, then the agent will be such that (iii) $\exists z(z \geq 0.8 \wedge ICr(\phi) = z)$ and (iv) $\exists z(z < 0.8 \wedge ICr(\phi) = z)$. That is if the agent meets the requirements imposed by QUANTITATIVE INDETERMINACY, then the agent will be such that it is indeterminate whether it believes ϕ . This suffices to show that, given reasonable background assumptions, INDETERMINACY falls out as a special case of QUANTITATIVE INDETERMINACY.

6.2 Supervaluationist Cognitive Significance

The situation with supervaluationist cognitive significance is for the most part the same as in the paracomplete case, with one major caveat. *Prima facie*, the most natural response to the attitudinal question is to accept REJECTION (now understood in supervaluationist terms) for the case of qualitative belief, and QUANTITATIVE REJECTION (again understood in supervaluationist terms) for the case of quantitative beliefs.

A natural way of representing an agent's credences while allowing for such constraints is in terms of Dempster-Shafer functions. This set of functions can be characterized as follows (where \models is understood in supervaluationist terms):

$$(D1) \models \phi \Rightarrow D_s(\phi) = 1$$

²³To see why any VF^+ agent will satisfy this constraint, it suffices to show that $Cr(D\phi \vee D\neg\phi) = x \models \forall y, z((Cr(D\phi) = y \wedge Cr(D\neg\phi) = z) \rightarrow y + z = x)$. To see that this holds assume: $\llbracket Cr(D\phi \vee D\neg\phi) = x \rrbracket = 1$. Then for every $Q \subseteq \Sigma^{\llbracket D\phi \vee D\neg\phi \rrbracket = 1/2}$ $\mathcal{P}(\Sigma^{\llbracket D\phi \vee D\neg\phi \rrbracket = 1} \cup Q) = x$. But the only way for this to hold is if $\mathcal{P}(\Sigma^{\llbracket D\phi \vee D\neg\phi \rrbracket = 1/2}) = 0$. So, at every ordinal stage α of the VF^+ model: (i) $\mathcal{P}(\Sigma_\alpha^{\llbracket D\phi \vee D\neg\phi \rrbracket = 1}) = x$, and $\mathcal{P}(\Sigma_\alpha^{\llbracket D\phi \rrbracket = 1/2}) = \mathcal{P}(\Sigma_\alpha^{\llbracket D\neg\phi \rrbracket = 1/2}) = 0$. These facts assure us that at every ordinal stage α of the VF^+ model $\mathcal{P}(\Sigma_\alpha^{\llbracket D\phi \rrbracket = 1} \cup \Sigma_\alpha^{\llbracket D\neg\phi \rrbracket = 1/2}) + \mathcal{P}(\Sigma_\alpha^{\llbracket D\neg\phi \rrbracket = 1} \cup \Sigma_\alpha^{\llbracket D\phi \rrbracket = 1/2}) = x$. This assures us that $\llbracket \forall y, z((Cr(D\phi) = y \wedge Cr(D\neg\phi) = z) \rightarrow y + z = x) \rrbracket = 1$.

$$(D2) \phi \models \perp \Rightarrow D_s(\phi) = 0$$

$$(D3) \phi \models \psi \Rightarrow D_s(\phi) \leq D_s(\psi)$$

$$(D4) D_s(\bigvee_{i=1}^m \phi_i) \geq \sum_S (-1)^{|S|-1} D_s(\bigwedge_{i \in S} \phi_i)^{24}$$

The salient feature of Dempster-Shafer functions is that the value assigned to a disjunction may be higher than the sum of the values assigned to the disjuncts. Since a supervaluationist will presumably hold that a rational agent should believe all supervaluationist theorems, such an agent should always assign credence 1 to $\phi \vee \neg\phi$. However, if ϕ is some proposition such that an agent should have credence 0.8 in the claim that it is indeterminate, then in order to satisfy the requirements imposed by QUANTITATIVE REJECTION, the agent would have to have credences in ϕ and $\neg\phi$ summing to at most 0.2. Such an agent's credence could not be represented by a probability function, but could be represented by a Dempster-Schafer function.²⁵

REJECTION and QUANTITATIVE REJECTION are both *prima facie* attractive accounts of the normative significance of the non-classical status of indeterminacy posited by supervaluationist theories. However, just as in the paracomplete case, we can argue that ultimately these principles are unacceptable. Given commitment to REJECTION and QUANTITATIVE REJECTION, we can resurrect the normative paradoxes developed in §§3.1-3.2, in a way that can't be resolved by appeal to supervaluationist resources.

I won't walk through the arguments. They are exactly the same as the arguments offered in §§6.1.1 - 6.1.2. Nothing in those arguments hinges on any feature that differs between the paracomplete and supervaluationist theories.

As in the paracomplete case, I suggest that we should, instead, accept INDETERMINACY as providing us with the correct account of the cognitive significance of (supervaluationist) indeterminacy in the case of qualitative beliefs. The proof of (24) will also go through for the supervaluationist case, *mutatis mutandis*; nothing in that argument hinges on any feature of VF^+ models that isn't also a feature of SV^+ models. And so, as in the paracomplete case, we can be assured that Alpha can meet all of the requirements imposed by EVIDENCE, CONSISTENCY and INDETERMINACY.

Unfortunately, for the case of quantitative beliefs, matters are less clear than in the paracomplete case. I'm still inclined to accept, in place of QUANTITATIVE REJECTION, QUANTITATIVE INDETERMINACY. However, unlike in the paracomplete case, in this case I don't know of any way of proving that Alpha can be guaranteed to meet all of the requirements imposed by the normative principles CREDENCE₁, CREDENCE₂, and QUANTITATIVE INDETERMINACY.

The problem is that our proof of (30) doesn't obviously go through once we switch to SV^+ models. In our earlier proof, appeal was made to a result, called the FUNDAMENTAL THEOREM. Unfortunately, the only proof that's been offered for this theorem no longer works when we switch from using VF^+ models to SV^+ models.²⁶ And I don't know of any alternative proof that will work for SV^+ models. Nor can I see how to prove (30) without appeal to this result. Now, I

²⁴To characterize the class of probability functions we need simply switch \geq to $=$ in (D4).

²⁵See Field (2000), for an endorsement of this way of thinking about the cognitive significance of supervaluationist non-classicality.

²⁶See Field (2008).

also don't know that the FUNDAMENTAL THEOREM *doesn't* hold for SV^+ models. Nor do I know that there aren't proofs of (30) that don't appeal to this result. At present, these are, so far as I know, open questions. And so, at present it is, so far as I know, an open question whether Alpha can be guaranteed to meet all of the requirements imposed by CREDENCE₁, CREDENCE₂ and QUANTITATIVE INDETERMINACY.

If we're supervaluationists, then, we should certainly reject QUANTITATIVE REJECTION. But whether we should replace this defective norm with QUANTITATIVE INDETERMINACY is something about which we should perhaps be somewhat hesitant. It should be noted, however, that there are some reasons for optimism here. For example, if we restrict our attention to propositions that only take either values 1, 1/2, or 0, i.e., propositions that have the same value at every ordinal stage of our SV^+ models, then it is easy to show that QUANTITATIVE INDETERMINACY will be satisfied by any SV^+ agent.²⁷ Since our normative paradoxes, at least as so far developed, have concerned only propositions in this fragment, then we get the suggestive result that no new normative paradoxes of this type would be instituted by endorsement of QUANTITATIVE INDETERMINACY in the supervaluationist case. Of course, this fragment of our language is rather small, and so this type of result is rather weak. But it nonetheless provides some reason to take QUANTITATIVE INDETERMINACY seriously as the answer to the attitudinal question in the supervaluationist case.

6.3 Paraconsistent Cognitive Significance

In the case of paracomplete and supervaluationist theories, we've labelled the statuses represented by semantic value 1/2 in our models *indeterminacy*. In the paraconsistent case, this label isn't really appropriate, since the paraconsistent theory *endorses* claims with this status. In this section, then, let's instead talk about a formula being *dialethic*. In place of our operator I , we will now employ an operator Δ . This is merely a notational change.²⁸

Now, as in the paracomplete and supervaluationist cases, we can ask about the cognitive significance of this non-classical status. In particular, we can ask: given that one ought to believe that a proposition ϕ is dialethic, what rational constraints does this impose on one's attitudes towards ϕ ? In the paraconsistent case (unlike in the paracomplete and supervaluationist cases) at least part of the answer to this question is clear. The following is, I think, something that any paraconsistent theorist will (and should) endorse:

ACCEPTANCE For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is dialethic, that one ought to believe both ϕ and its negation.²⁹

²⁷It's rather easy to adopt the paracomplete proof of QUANTITATIVE INDETERMINACY to the supervaluationist case, given this simplifying assumption. I leave this to the interested reader.

²⁸We say $\Delta\phi =_{df} (\neg\phi \vee (\phi \rightarrow \neg\phi)) \wedge (\phi \vee (\neg\phi \rightarrow \phi))$. This is, of course, just the definition of the indeterminacy operator I . I should note that this is in some ways a more restricted use of the term 'dialethic' than is traditional. Most authors take 'dialethic' to just mean true contradiction. This is a more general notion, since there are other designated values that a contradiction could have besides value 1/2. I trust, though, that, this being noted, it won't cause any confusion.

²⁹Read: $OB\Delta\phi \models OB\phi \wedge OB\neg\phi$.

I say that a paraconsistent theorist should accept ACCEPTANCE. Why? Well, semantic value $1/2$ is a *designated* value according our paraconsistent theory. If a sentence has semantic value $1/2$ in every model, then it is a theorem. Plausibly, one should believe all theorems, and so one should believe any claim that is guaranteed to have semantic value $1/2$. That is, one should believe any claim that is guaranteed to be dialethic. At the very least, if one ought to believe that ϕ is a theorem, then one ought to believe ϕ . And so, at the very least, if one ought to believe that ϕ is guaranteed to be dialethic, then one ought to believe ϕ . Of course, this isn't quite the claim that if one ought to believe that ϕ is dialethic, then one should believe it. But I can't see how one could accept the former without admitting the latter. To get the claim that one should also believe its negation, we need simply note that ϕ will be dialethic just in case $\neg\phi$ is dialethic.

I think that ACCEPTANCE provides a partial answer to the attitudinal question as it applies to a paraconsistent theory. But why only partial? Well, as I argued in §4.1.4 and §5.1, if one is a paraconsistent theorist, then one should allow that claims about an agent's beliefs may be dialethic. If an agent ought to believe that ϕ is dialethic, then, by ACCEPTANCE, the agent ought to believe ϕ and $\neg\phi$. Assuming that agent meets these requirements, it may, nonetheless, still be that the agent fails to believe ϕ or fails to believe $\neg\phi$; this is just to sort of weirdness that we need to put up with, if we buy into a paraconsistent theory. One thing that we want to know, then, is what would the normative status of failing to believe ϕ , or $\neg\phi$, be in this situation?

It might seem that the answer to this question is obvious. Given ACCEPTANCE, we know that if one ought to believe that ϕ is dialethic, then one ought believe ϕ and believe $\neg\phi$. But from this, the following general principle might seem to follow:

PROHIBITION For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is dialethic, that both not believing ϕ and not believing $\neg\phi$ are prohibited.³⁰

Here's why. *Prima facie* the following is plausible :

(P-O) ϕ is prohibited just in case $\neg\phi$ is obligatory.³¹

Since by ACCEPTANCE, we have that it follows from $OB\Delta\phi$ that $OB\phi$ and $OB\neg\phi$, if we accept (P-O), then we will hold that it also follows from $OB\Delta\phi$ that $P\neg B\phi$ and $P\neg B\neg\phi$. That is, given ACCEPTANCE and (P-O), we have PROHIBITION.

Now, in a classical setting (or even in a paracomplete or supervaluationist setting) (P-O) seems to me to be fairly reasonable. However, in a paraconsistent setting, the principle seems to me to be highly non-obvious. If ϕ is prohibited, then one will be subject to rational criticism if ϕ is realized. Now the question we want to ask is: if $\neg\phi$ is obligatory, need one be subject to rational criticism, if one realizes ϕ ? Given paraconsistent commitments, it isn't obvious that the answer is 'yes'. For, if, in addition to realizing ϕ , one realizes $\neg\phi$, one will thereby have met one's obligation. In such cases, it seems to me that rational criticism may be inappropriate. In order to allow for this, we need to distinguish between ϕ being prohibited and and its negation being obligatory.

³⁰Read: $OB\Delta\phi \models P\neg B\phi \wedge P\neg B\neg\phi$.

³¹Read: $P\phi \leftrightarrow O\neg\phi$

Indeed, I think that reflection on some of our preceding arguments provide us with excellent reason to insist on this distinction. For, unless we allow for this distinction, we will once again be saddled with a normative paradox.

In §4.1.4, I argued that a paraconsistent theorist can resolve the normative paradox developed in §3.1, but doing so required saying that Alpha must both believe that $\lceil\beta\rceil$ is true and not believe that $\lceil\beta\rceil$ is true. In this case, then, it is rationally obligatory that Alpha believe that $\lceil\beta\rceil$ is true, and rationally obligatory that Alpha not believe that $\lceil\beta\rceil$ is true. If, however, we endorse (P-O), and so say that an agent is subject to rational criticism for realizing some option ϕ just in case it was rationally obligatory that the agent realize $\neg\phi$, then we will be forced to say that in this situation Alpha will be subject to rational criticism no matter what it does. For whether Alpha believes that $\lceil\beta\rceil$ is true, or doesn't believe that $\lceil\beta\rceil$ is true, it will be doing something impermissible. Since I think we should take POSSIBILITY as a basic constraint on principles of normativity, i.e., we should hold that an agent can always in principle avoid being subject to rational criticism, I think that we should reject (P-O). We shouldn't think that rational prohibition can be analyzed in terms of rational obligation.³²

To see why this sort of view isn't all that implausible (at least in a paraconsistent setting), it helps to think of rational obligations as setting conditions under which an agent may be subject to *rational approbation*, and to think of rational prohibitions as setting conditions under which an agent may be subject to *rational criticism*. Seen this way, (P-O) amounts to the claim that an agent is appropriately subject to rational criticism for realizing ϕ just in case the agent is appropriately subject to rational approbation for realizing $\neg\phi$. I don't think that this is really at all obvious in a paraconsistent setting. In such a setting, it seems plausible that the conditions for appropriate criticism may have to be stipulated separately from the conditions for appropriate approbation. And, indeed, as we've seen, there are good theoretical reasons to think that, given paraconsistent background assumptions, this the case; otherwise, certain agents may have no way of escaping rational criticism. We should allow, then, that in certain circumstances an agent may be subject to rational approbation for realizing $\neg\phi$, without thereby being subject to rational criticism for realizing ϕ .

Since I don't think that we should accept (P-O), we shouldn't, then, accept the argument just canvassed from ACCEPTANCE to PROHIBITION. This negative point, of course, doesn't answer the question: should we accept PROHIBITION? But the answer, I claim, is 'no'. Once again, we can marshal our results from earlier chapters in defense of this claim. Acceptance of PROHIBITION leads us back in to normative paradox.

In order to resolve the normative paradox developed in §3.1, a paraconsistent theorist must hold that Alpha ought to be such that it both believes and doesn't believe that $\lceil\beta\rceil$ is true. I take it that a rational way for Alpha to meet this obligation is for Alpha to be such that it is dialethic

³²One way that a paraconsistent theorist might try to resist this would be to allow normative claims to be sources of dialethias. In this way we could say that such an agent would not be subject to rational criticism (despite their also being subject to rational criticism). I'm inclined, however, to think that we should take POSSIBILITY to provide a stricter constraint on rational principles, one that can't be met in this dialethic manner. In what follows, then, I'll ignore this option, although I think it is at least a coherent, if not terribly attractive, position for a paraconsistent theorist.

that it believes that $\ulcorner\beta\urcorner$ is true.³³ Given that we have $\Delta B_\alpha T(\ulcorner\beta\urcorner)$ we also have $\Delta\neg B_\alpha T(\ulcorner\beta\urcorner)$. And since we have $\models T(\ulcorner\beta\urcorner) \leftrightarrow \neg B_\alpha T(\ulcorner\beta\urcorner)$, it follows from $\Delta\neg B_\alpha T(\ulcorner\beta\urcorner)$ that $\Delta T(\ulcorner\beta\urcorner)$. We'll assume, further, that Alpha is perfectly reliable about whether or not the claim $\ulcorner\beta\urcorner$ is true is dialethic. That is, I assume:

$$(32) B_\alpha \Delta T(\ulcorner\beta\urcorner) \leftrightarrow \Delta T(\ulcorner\beta\urcorner)$$

This claim will hold in any of the VF^+ models in which we've represented Alpha's doxastic state, so it shouldn't be particularly controversial.

Given, then, our assumption that Alpha is rationally such that it is dialethic that it believes that $\ulcorner\beta\urcorner$ is true, it follows that it is dialethic that $\ulcorner\beta\urcorner$ is true, and so that Alpha believes that it is dialethic that $\ulcorner\beta\urcorner$ is true. I take it that Alpha is not irrational simply in virtue of the reliability ascribed by (32). Given acceptance of PROHIBITION, it then follows that Alpha is prohibited from not believing the $\ulcorner\beta\urcorner$ is true, i.e., Alpha will be subject to rational criticism if it is such that it does not believe that $\ulcorner\beta\urcorner$ is true. But, as we established earlier, the only way for Alpha to avoid being subject to rational criticism, given the paradox developed in §3.1, is if Alpha doesn't believe that $\ulcorner\beta\urcorner$ is true. It follows, then, that if we accept PROHIBITION, we will be forced to admit that Alpha, an antecedently rational agent, has no option available to it that will allow it to avoid incurring rational criticism. Since, I think we should always allow that a rational agent can be such that it can avoid rational criticism, we should reject PROHIBITION.

In some cases, then, we should allow that a rational agent may believe that ϕ is dialethic and fail to believe ϕ without incurring rational criticism. Of course, given ACCEPTANCE, such an agent will also believe ϕ . But as I've argued, there is good reason for a paraconsistent theorist to allow for these types of paraconsistent doxastic states.

We shouldn't, then, accept that *in general* if an agent ought to believe that ϕ is dialethic, then the agent is prohibited from not believing ϕ and prohibited from not believing $\neg\phi$. Might we nonetheless hold that in some, perhaps most, cases such prohibitions follow from the rational obligation to believe that ϕ is dialethic?

So far as I can see, this would be a consistent position. But as with similarly hedged positions in the paracomplete and supervaluationist cases, I can't see any principled reason for holding this view. We can't hold that such prohibitions are present in the case of the proposition expressed by $\ulcorner\beta\urcorner$. The question, then, is why such prohibitions are present, say, in the case of the proposition expressed by the liar sentence $\ulcorner\lambda\urcorner$, but not in the case of the proposition expressed by $\ulcorner\beta\urcorner$? And here I can't see what a satisfactory answer might look like. Simply making an exception in the case of $\ulcorner\beta\urcorner$, in order to avoid normative paradox, seems to me completely *ad hoc*. Lacking any principled grounds, then, for distinguishing between cases like $\ulcorner\lambda\urcorner$ and cases like $\ulcorner\beta\urcorner$, we should, I think, treat the normative significance of the dialethic status of the propositions expressed by these sentences in the same manner. Since we should reject PROHIBITION in the case of $\ulcorner\beta\urcorner$, we should reject other instances of PROHIBITION as well.

³³Given that we are using 'dialethic' as term for the status that is modeled by semantic value 1/2 in the VF^+ models employed in our paraconsistent theory, there will be other statuses that $B_\alpha T(\ulcorner\beta\urcorner)$ could have that would allow Alpha to meet the above normative demands. But as in the paracomplete and supervaluationist cases, I can't see any principled reason for denying that this particular status would be a rational way to meet the relevant obligations.

Our reflections so far, then, have led us to conclude that we should reject PROHIBITION and instead accept the following alternative normative principle connecting belief that ϕ is dialethic and failure to believe ϕ and $\neg\phi$:

PERMISSION For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is dialethic, that both not believing ϕ and not believing $\neg\phi$ are permitted.³⁴

A natural question, at this point, is whether, in addition to such failures of belief being permitted, they might, in addition, be required? That is, a natural question is whether in addition to PERMISSION we should accept the following:

NEGATIVE BELIEF For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is dialethic, that one ought to not believe ϕ and one ought to not believe $\neg\phi$.³⁵

The answer I want to (tentatively) suggest is ‘yes’. Indeed, I’m inclined to think that we should accept the following principle:

DIALETHIC BELIEF For any proposition ϕ , it is a consequence of the claim that one ought to believe that ϕ is dialethic, that one ought to be such that it is dialethic that one believes ϕ .³⁶

From this principle, both ACCEPTANCE and NEGATIVE BELIEF can be shown to follow:

DIALETHIC BELIEF entails ACCEPTANCE

First we’ll show that $OB\Delta\phi \models OB\phi$. Assume $OB\Delta\phi$. By DIALETHIC BELIEF, we have $O\Delta B\phi$. We also have, $\Delta B\phi \models B\phi$, and since, I take it, rational obligations are closed under consequence, we then have $OB\phi$. Next we’ll show that $OB\Delta\phi \models OB\neg\phi$. Again assume $OB\Delta\phi$. We have $\models \Delta\phi \leftrightarrow \Delta\neg\phi$. Given this and the fact that we have $OB\Delta\phi$, we should infer $OB\Delta\neg\phi$. (One way of arguing for this is by appealing to the idea that one ought to believe theorems. This, together with the general principle that oughts are closed under consequence would let us infer $OB\Delta\neg\phi$ from $OB\Delta\phi$.) By DIALETHIC BELIEF, it follows that we have $O\Delta B\neg\phi$. And since we have $\Delta B\neg\phi \models B\neg\phi$, it follows that we have $OB\neg\phi$.

³⁴Where: Permitted $\phi =_{df} \neg$ Prohibited ϕ . We should then read the above principle as saying: $OB\Delta\phi \models \neg P\neg B\phi \wedge \neg P\neg B\neg\phi$.

³⁵Read: $OB\Delta\phi \models O\neg B\phi \wedge O\neg B\neg\phi$. This type of principle is considered in Priest (2006a), although it is not unequivocally endorsed. Priest worries that it will result in a rational dilemma, since this principle combined with ACCEPTANCE results in the demand that an agent both believe and not believe certain propositions, and this, Priest thinks, is not possible. This isn’t a worry, however, that I take particularly seriously, since I think we have good reason, if we’re paraconsistent theorists, to allow for dialethic belief states.

³⁶Read: $OB\Delta\phi \models O\Delta B\phi$.

DIALETHIC BELIEF entails NEGATIVE BELIEF

First we'll show that $OB\Delta\phi \models O\neg B\phi$. The argument is essentially the same as the first case above. Given that we have $OB\Delta\phi$, by DIALETHIC BELIEF, we have $O\Delta B\phi$. And since we have $\Delta B\phi \models \neg B\phi$, it follows from the closure condition on rational obligations that $O\neg B\phi$. Next we'll show that $OB\Delta\phi \models O\neg B\phi$. Again the case quite similar to the second case above. There we showed that it follows from $OB\Delta\phi$ that $O\Delta B\neg\phi$. And since we have $\Delta B\neg\phi \models \neg B\neg\phi$, and since obligations are closed under consequence, we have $O\neg B\neg\phi$.

Why should we accept DIALETHIC BELIEF? The principle is obviously analogous to INDETERMINACY. However, the status of DIALETHIC BELIEF for a paraconsistent theory may seem less clear than the status of INDETERMINACY for a paracomplete theory. In the latter case, INDETERMINACY seemed to be *the only* plausible principle connecting the normative status of one's attitudes towards an indeterminate proposition ϕ and one's attitudes towards ϕ . In paraconsistent case, however, we've already seen that there are at least two plausible such principles, viz., ACCEPTANCE and PERMISSION. Is there, then, good reason to accept, in addition to these principles, DIALETHIC BELIEF? I think there is.

The following is an attractive thought. If an agent is perfectly rational, then her doxastic state should be able to be modeled by a set of logically possible worlds \mathcal{W} , such that the agent will believe a proposition ϕ just in case ϕ is true at all $w \in \mathcal{W}$.³⁷ Now the important point for our purposes is that if we model an agent's doxastic state in this way, and we're paraconsistent theorists, then the agent will be guaranteed to meet the requirements imposed by DIALETHIC BELIEF.³⁸ An idealized rational agent's doxastic states, then, will follow the patterns imposed by DIALETHIC BELIEF. This gives us good reason, I think, to accept DIALETHIC BELIEF.

I note, also, that this gives us good reason, if we are paraconsistent theorists, to accept CONSISTENCY. For, as we saw in §4.1.2, any agent whose doxastic state is representable in this way will satisfy the demands imposed by this principle.

This same line of thought also gives the paraconsistent theorist good reason to accept the following constraint on credences:

QUANTITATIVE DIALETHIC BELIEF For any proposition ϕ , it is a consequence of the claim that one ought to be such that $Cr(D\phi) = x$ and $Cr(D\neg\phi) = z$, that one ought to be such that the following all hold:

- (i) $\forall y(y < x \rightarrow D(Cr(\phi) \neq y))$
- (ii) $\forall y(y > 1 - z \rightarrow D(Cr(\phi) \neq y))$

³⁷If one is worried that rationality shouldn't demand that one believe certain necessary truths, such as water = H₂O, then one can allow that the logically possible worlds exceed the metaphysically possible. This thought, however, does mean that a perfectly rational agent will believe all logical truths. But who said that being rational was easy.

³⁸The proof of this claim is the same as the proof, given in §6.1.1, that any VF^+ agent will meet the requirements imposed by INDETERMINACY.

$$(iii) x \neq 1 - z \rightarrow (\Delta Cr(\phi) = x \wedge \Delta Cr(\phi) = 1 - z)$$

$$(iv) x = 1 - z \rightarrow (DCr(\phi) = x)$$

For just as it's plausible to think that an ideally rational agent's qualitative belief state can be represented by set of possible worlds, it's also plausible that such an agent's quantitative belief state can be represented by a probability measure over a space of possible worlds.³⁹ But as we've already seen, if we think of possible worlds in paraconsistent terms, then any agent whose credal state can be represented in this way will meet the requirements imposed by QUANTITATIVE DIALETHIC BELIEF.⁴⁰

I note that this also gives us an argument for the claim that a paraconsistent theorist should accept CREDENCE₂. For, as we saw in §4.2.1, if we think of possible worlds in paraconsistent terms, then an agent whose credal state can be represented in this way will also meet the requirements imposed by this norm.

These considerations are certainly not completely decisive. But I think they give good reason to take seriously the idea that in the paraconsistent case, as in the paracomplete and supervaluationist cases, a rational agent's response to a paradoxical proposition such as that expressed by $\ulcorner \lambda \urcorner$ should be to have doxastic attitudes towards that proposition whose status mirrors the paradoxical status of the proposition in question. In the paraconsistent case, this means that in responding to dialethic propositions an agent will have dialethic beliefs and dialethic credences.

³⁹There's actually some tension between these two ideas, at least if one thinks that qualitative belief should be understood in terms of one having a credence at or above some threshold $\tau < 1$. For, if one thinks that an agent's qualitative belief state can be represented by a set of possible worlds, then such an agent's beliefs will be closed under logical consequence. But the same will not be true if you think that such an agent's credences can be represented by a probability measure, and, in addition, think that belief is just credence at or above some threshold less than 1. For our purposes, however, we need not worry too much about this tension. For DIALETHIC BELIEF will be a consequence of QUANTITATIVE DIALETHIC BELIEF. Thus, even if, say, we ultimately want to abandon the thought that qualitative belief states can be represented by sets of possible worlds, this need not lead us to reject DIALETHIC BELIEF.

⁴⁰This follows, given claims (28) - (31) in §6.1.

Bibliography

- Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis*, 68:277–297.
- Boolos, G. and Jeffrey, R. (1989). *Computability and Logic*. Cambridge University Press.
- Burge, T. (1978). Buridan and epistemic paradox. *Philosophical Studies*, 34:21–35.
- Burge, T. (1984). Epistemic paradox. *Journal of Philosophy*, 81(1):5–29.
- Conee, E. (1982). Against moral dilemmas. *The Philosophical Review*, 91(1):87–97.
- Conee, E. (1987). Evident, but rationally unacceptable. *Australasian Journal of Philosophy*, 65:316–326.
- Davidson, D. (1980a). Actions, reasons, causes. In *Essays on Actions and Events*. Oxford University Press.
- Davidson, D. (1980b). Mental events. In *Essays on Actions and Events*. Oxford University Press.
- de Finetti, B. (1974). *Theory of Probability*, volume 1. John Wiley and Sons.
- Elga, A. and Egan, A. (2005). I can't believe I'm stupid. *Philosophical Perspectives*, 19.
- Feldman, R. (2004). Having evidence. In *Evidentialism*, chapter 9. Oxford University Press.
- Field, H. (2000). Indeterminacy, degree of belief and excluded middle. *Nous*, 34:1–30.
- Field, H. (2003). No fact of the matter. *Australasian Journal of Philosophy*, 81(4):457–480.
- Field, H. (2007). Solving the paradoxes, escaping revenge. In Beall, J. C., editor, *Revenge of the Liar*. Oxford University Press.
- Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.
- Fodor, J. (1975). *The Language of Thought*. Harvard University Press.
- Friedman, H. and Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21.
- Gibbard, A. (2008). Rational credence and the value of truth. In *Oxford Studies in Epistemology*, volume 2. Oxford University Press.
- Gibbard, A. and Harper, W. (1978). Counterfactuals and two kinds of expected utility. In Leach, C., McClennen, E., and Hooker, C., editors, *Foundations and Applications of Decision Theory*, pages 125–162. Dordrecht: D. Reidel.
- Gupta, A. and Belnap, N. (1993). *The Revision Theory of Truth*. MIT Press.
- Halbach, V. (2011). *Axiomatic Theories of Truth*. Cambridge University Press.
- Hintikka, J. (1962). *Knowledge and Belief*. Cornell University Press.
- Horwich, P. (1982). *Probability and Evidence*. Cambridge University Press.
- Jeffrey, R. (1983). *The Logic of Decision*. Chicago University Press, 2 edition.
- Joyce, J. (1998). A non-pragmatic vindication of probabilism. *Philosophy of Science*, 65:575–603.

- Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.
- Joyce, J. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*. Synthese Library.
- Kaplan, D. and Montague, R. (1960). A paradox regained. *Notre Dame Journal of Formal Logic*, 1(3):79–90.
- Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716.
- Leitgeb, H. and Pettigrew, R. (2010a). An objective justification of bayesianism I: Measuring inaccuracy. *Philosophy of Science*, 77.
- Leitgeb, H. and Pettigrew, R. (2010b). An objective justification of bayesianism II: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77.
- Lemmon, E. (1962). Moral dilemmas. *The Philosophical Review*, 71(2).
- Lewis, D. (1974). Radical interpretation. *Synthese*, 23:331–44.
- Lewis, D. (1986). A subjectivist’s guide to objective chance. In *Philosophical Papers Vol.II*. Oxford University Press.
- Lewis, D. (1999). Reduction of mind. In *Papers on Metaphysics and Epistemology*. Cambridge University Press.
- Lewis, D. (2000). Logic for equivocators. In *Papers in Philosophical Logic*. Cambridge University Press.
- Maher, P. (1993). *Betting on Theories*. Cambridge University Press.
- Makinson, D. (1965). The paradox of the preface. *Analysis*, 25(6):205–207.
- Marcus, R. B. (1980). Moral dilemmas and consistency. *Journal of Philosophy*, 77(3):121–136.
- Maudlin, T. (2004). *Truth and Paradox*. Oxford University Press.
- McGee, V. (1991). *Truth, Vagueness and Paradox*. Hackett Publishing Company.
- Osborne, M. and Rubenstein, A. (1994). *A Course in Game Theory*. MIT Press.
- Parsons, T. (1984). Assertion, denial and the liar paradox. *Journal of Philosophical Logic*, 13:137–152.
- Priest, G. (2002). Rational dilemmas. *Analysis*, 62(1):11–16.
- Priest, G. (2006a). *Doubt Truth To Be A Liar*. Oxford University Press.
- Priest, G. (2006b). In *Contradiction*. Oxford University Press, 2 edition.
- Ramsey, F. (1931). Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*. Kegan Paul.
- Richard, M. (2008). *When Truth Gives Out*. Oxford University Press.
- Savage, L. (1972). *The Foundations of Statistics*. Dover Publications.
- Soames, S. (1999). *Understanding Truth*. Oxford University Press.
- Sorensen, R. (1988). *Blindspots*. Oxford University Press.
- Stalnaker, R. (1984). *Inquiry*. MIT Press.
- Tarski, A. (1956). The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics*. Oxford University Press.
- Van Fraassen, B. C. (1973). Values and the heart’s command. *Journal of Philosophy*, 70(1).
- Wedgwood, R. (2008). Gandalf’s solution to the newcomb problem. (MS).
- Williamson, T. (1994). *Vagueness*. Routledge.
- Williamson, T. (2000). *Knowledge and it’s Limits*. Oxford University Press.

-
- Yablo, S. (2003). New grounds for naive truth theory. In Beall, J. C., editor, *Liars and Heaps*. Oxford University Press.