Belief and Indeterminacy

Michael Caie

§1 Introduction

What attitude should a rational agent have towards a proposition expressed by a paradoxical sentence such as the Liar? In this paper, I argue that consideration of a paradox concerning doxastic rationality motivates a surprising answer to this question.

The question I'm interested in is not the same as the question about what semantic status(es) we should assign to such paradoxical sentences. But the two questions are intimately related. Sometimes, given an account answering the latter question, the answer to the former is obvious.[1] But this is not the case for every account of semantic paradox. For the type of account that I'll be interested in here, the answer to our question is far from obvious.

This account traces back to Kripke (1975).[2] According to the account, ϕ and $T\ulcorner\phi\urcorner$ are intersubstitutable (at least within extensional contexts). In order to achieve this desirable goal, however, certain elements of classical logic must be abandoned. According to this theory, the logic governing the "Boolean" connectives and quantifiers is K_3, the logic induced by the Strong Kleene valuation scheme.[3] Notably, in this logic excluded-middle fails to be unrestrictedly valid. Such theories are labelled *paracomplete*.

This type of account has been extended by Hartry Field. Field has provided a model-theory

[1]For example, some theories hold that the Liar sentence is not true. But this, of course, is just what the Liar sentence says. So the proponent of such a theory will hold that one should believe the proposition expressed by the Liar sentence.

[2]Kripke's paper has given rise to a number of theories. The relevant theory for our purposes is called 'KFS'. See also Soames (1999), Field (2008), Richard (2008), Yablo (2003).

[3]See the Technical Appendix for the details of this valuation.

which, like Kripke's, allows for the intersubstitutability of ϕ and $T\ulcorner\phi\urcorner$ by allowing for the failure of excluded-middle.[4] The models developed by Field, however, are defined for a language containing a reasonable conditional, '$\rightarrow$', for which the T-schema holds, and an indeterminacy operator, 'I'. Using the latter we can characterize the status of those paradoxical sentences that lead to failures of excluded-middle. While we cannot, without committing ourselves to a contradiction, say either that the Liar sentence λ or its negation are not true, we can say that neither λ nor its negation are *determinately* true, i.e., we can say that it is *indeterminate* whether λ is true.

The arguments that follow are directed at anyone who advocates a paracomplete treatment of the semantic paradoxes. In what follows, I will, however, help myself to the expressive resources that Field's theory offers. Because Field's theory contains an object language operator 'I', which can be used to express the paradoxical status of the Liar sentence, working with this theory will allow us to reason more easily about this status and our attitudes towards sentences having this status. In principle, however, the main arguments that follow could be reframed in terms acceptable to advocates of alternative paracomplete accounts.

Letting ϕ be some proposition that one ought to believe is indeterminate, the question then arises for one who is attracted to this sort of account of semantic paradox: What attitude should one have towards ϕ? Call this *The Normative 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, i.e., disbelieve, both ϕ and its negation.[5]

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

[4] See Field (2008), Field (2007) for a development of this position.

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

the proposition expressed by the Liar sentence.[6]

In what follows, I'll argue, instead, that the correct answer to The Normative 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.[7]

For rational agents, indeterminacy in the objects of doxastic states will filter up to the doxastic states themselves. This is in many ways a strange and counterintuitive claim. Nonetheless, I'll try to show that there are good reasons to think it true.

The argument for this will proceed as follows.

In §2.1, I develop a normative paradox. I show that three plausible principles concerning doxastic rationality are classically inconsistent.

In §2.2, I show how the *prima facie* inconsistent triad can be whittled down to a *prima facie* inconsistent pair.

In §3, I show that the structure of the paradox developed in §2.1 is formally identical to a modal Liar paradox. A paracomplete solution to the Liar paradox can be straightforwardly extended to deal with the modal Liar sentence. Given the structural identity of these two paradoxes, the normative paradox can also be resolved by appeal to indeterminacy. The same solution applies to the paradox developed in §2.2.

The possibility of doxastic indeterminacy is defended in §4.

In §5, I show that resolving the normative paradoxes in this way requires rejecting REJECTION. Acceptance of this norm leads us back into normative paradox. By appealing to indeterminacy we can, it seems, hold on to some very plausible principles concerning doxastic rationality. But only if we give up REJECTION. Whatever confidence we have in these principles should make us correspondingly less confident in REJECTION. Since these principles are much more intuitively compelling than REJECTION, this provides a strong argument against REJECTION.

[6]See Field (2008) and Field (2003a) 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).

[7]Read: $OBI\phi \models OIB\phi$.

In §6, I argue that INDETERMINACY is independently motivated and that, unlike REJECTION, it is compatible with the proposed resolution of the normative paradox.

In §7, I assess the state of the dialectic between the proponent of REJECTION and the proponent of INDETERMINACY, given the preceding arguments.

Finally, in §8, I consider three *prima facie* compelling arguments in favor of REJECTION and show how they can be resisted.[8]

§2 Normative Paradoxes

§2.1

Consider the following sentence:

> **I do not believe that this sentence is true**[9]

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.

[8]Included is also an appendix that outlines the underlying technical machinery and proves various results that I appeal to in the main body of the paper. Readers who wish to skip this will not, however, miss anything essential to understanding the main arguments of the paper.

[9]This type of sentence and some of its odd features are discussed in Burge (1978), Burge (1984), Conee (1987) and Sorensen (1988). The paradox I develop, however, differs from those discussed by these authors in some important details.

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 $\neg\phi$.[10]

EVIDENCE For any proposition ϕ, if an agent's evidence makes ϕ certain then the agent is rationally required to believe ϕ.[11]

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 Given a set of mutually exclusive and jointly exhaustive doxastic options there is always some option such that it is possible for an agent, who is not already guilty of a rational failing, to realize that option and not incur rational criticism in so doing.

Later in this paper, I will show how this paradox can be resolved, and what lessons we can extract from its resolution. But first let us see how this paradox works in detail.

Let 'B_α' be an operator meaning *Alpha believes that*. Let 'β' name the following sentence: '$\neg B_\alpha T(\beta)$'.[12] Then, as an instance of the T-schema, we have:

[10]Read: $O(B\phi \rightarrow \neg B\neg\phi)$.

[11]Two points. (i) N.B. 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.

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

(1) $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$[13]

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 ϕ, then it also believes that it believes ϕ. And, whenever Alpha does not believe ϕ, then it believes that it does not believe ϕ. Moreover, Alpha is, overall, *perfectly* reliable in the higher-order beliefs that it has. Alpha believes that it believes ϕ only if it does believe ϕ, and it believes that it does not believe ϕ only if it does not believe ϕ.

We have, then, the following:

(2) $B_\alpha T(\beta) \leftrightarrow B_\alpha B_\alpha T(\beta)$

(3) $\neg B_\alpha T(\beta) \leftrightarrow B_\alpha \neg B_\alpha T(\beta)$[14]

We may further suppose that our agent believes the truth expressed in (1). We have then:

(4) $B_\alpha(B_\alpha T(\beta) \rightarrow \neg T(\beta))$

[13]N.B. '$\leftrightarrow$' is the Field biconditional. All of the inferences involving the conditional that I'll be appealing to are valid in the class of Field models. See the Technical Appendix for an outline of this model theory and Field (2008) for a more in depth treatment.

[14]We should think of 'B_α' as involving a first-personal mode of presentation for Alpha.

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

The possibility of an agent, such as Alpha, who satisfies (2)-(5), 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 β is true. By (3), it follows that it believes that it does not believe this. Alpha also believes that if it does not believe this then β is true. This is (5). The set-up of the case is such that the status of both of these beliefs is superlative. The first belief is, *qua* higher-order belief, *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.[15] We can assume, further, that Alpha has no other evidence that bears one way or the other on the question of whether β is true. Alpha, then, would seem to be in the position in which its evidence makes it certain that β is true. By EVIDENCE, it follows that Alpha ought to believe that β is true.

[15]Exactly 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 β is true, and that if it does not believe that β is true then β 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 e.g., 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, e.g., having a justified belief in ϕ. See e.g., Feldman (2004).

Case 2: On the assumption that Alpha does believe that β is true, it follows that it ought not believe that β is true.

> Assume that Alpha does believe that β is true. By (2), it follows that it believes that it believes that β is true. Alpha also believes that if it believes that β is true then β is not true. This is (4). 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 β is true. Alpha, then, is such that its evidence makes it certain that β is not true. By EVIDENCE it follows that Alpha ought to believe that β is not true. That is, we have $OB_\alpha\neg T(\beta)$. CONSISTENCY tells us $O(B_\alpha T(\beta) \rightarrow \neg B_\alpha\neg T(\beta))$. $B_\alpha T(\beta) \rightarrow \neg B_\alpha\neg T(\beta)$ is logically equivalent to $B_\alpha\neg T(\beta) \rightarrow \neg B_\alpha T(\beta)$. 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.[16] Given this it follows from CONSISTENCY that $O(B_\alpha\neg T(\beta) \rightarrow \neg B_\alpha T(\beta))$. From this, $OB_\alpha\neg T(\beta)$, and the following instance of the closure principle, $O(B_\alpha\neg T(\beta) \rightarrow \neg B_\alpha T(\beta)), OB_\alpha\neg T(\beta) \models O\neg B_\alpha T(\beta)$, we then have $O\neg B_\alpha T(\beta)$.

Cases 1 and 2 show that CONSISTENCY and EVIDENCE are jointly inconsistent with POSSIBILITY. If Alpha does not believes that β is true, then, according to Case 1, it ought to believe it. While if Alpha does believe that β is true, then, according to Case 2, it ought to not believe it. It follows that, given CONSISTENCY and EVIDENCE, there is a set of mutually exclusive, jointly exhaustive, doxastic options no member of which can be realized without incurring rational criticism.[17]

[16] This sort of multi-premise closure principle is not completely uncontroversial. In particular those who think that there are rational dilemmas, i.e., cases in which $O\phi$ and $O\neg\phi$ will be inclined to deny this, since typically they will deny that it is ever true that $O(\phi \wedge \neg\phi)$. E.g., see 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 the consequent is logically possible. 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.

[17] Note 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, viz., that (2), (3), (4) and (5) 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.

Nothing, moreover, about the set-up of the case would seem to require that Alpha be initially guilty of any rational failing. All that we require is that Alpha have knowledge of a theorem and that it be sensitive to its own doxastic states. The case of Alpha thus leads to a violation of POSSIBILITY.

While CONSISTENCY, EVIDENCE and POSSIBILITY appear to be jointly inconsistent, individually each is extremely plausible. CONSISTENCY seems to me to be as close to a non-negotiable principle of rationality as any we have, while EVIDENCE follows from the natural idea that a rational agent will be responsive to its evidence. It would be quite a cost, I think, if we were forced to give up one of these principles. Perhaps, then, one might think that the lesson to draw is that we should reject POSSIBILITY. We should not, however, underestimate the intuitive costs of this response. *Prima facie* it is, I think, 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.[18]

This intuition can be bolstered by considering the sorts of conditions under which rational criticism seems to be appropriate. 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.

[18]I should note that not everyone is inclined to accept this. There are those who think that moral and rational dilemmas—cases in which you are damned if do and damned if you don't—may arise even if an agent is not already subject to such criticism. See e.g., Lemmon (1962) and Marcus (1980) for the existence of moral dilemmas. For the existence of rational dilemmas see e.g., the case of Death in Damascus in Gibbard and Harper (1978), and Priest (2002). For resistance to the idea of rational and moral dilemmas see e.g., Conee (1982) and Arntzenius (2008). I do not here have the space to deal with this substantial literature, but it suffices to say my sympathies are with those who want to reject the possibility of such cases. As a minimal point, let me note that all parties to the dispute should, I think, agree that it would be ideal if we could find a way for the putative normative principles to coexist without conflict. In game-theory, a standard move to deal with putative cases of rational dilemmas is to expand the space of possible options and allow for so-called mixed- decisions. See e.g., Osborne and Rubenstein (1994). To foreshadow somewhat, my plan is to offer a similar strategy for the doxastic case. It is, I hope, enough to motivate this response that one sees the pull of POSSIBILITY.

(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.

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 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. If one wants to maintain what is, I think, a natural principle about the conditions for appropriate rational criticism then one should endorse POSSIBILITY.[19]

[19] Here's the argument for this claim: Let Γ pick out the set of doxastic options that a rationally blameless agent has realized. Let Σ be a set of mutually exclusive and jointly exhaustive options 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., $\Gamma \cup \Gamma' \cup \sigma \models \bot$. It follows that $\Gamma \cup \Gamma' \models \neg\sigma$. (N.B. $\{\sigma, \neg\sigma\}$ are, by the set-up of the case, exhaustive, and so this particular use of *reductio* is fine by both classical and paracomplete lights.) 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 $\Sigma - \sigma$.

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.

§2.2

In this section, I provide a different case which whittles the apparent inconsistency down to the pair CONSISTENCY and POSSIBILITY. Although I'm not inclined to reject EVIDENCE, I can imagine that if one felt forced to choose between CONSISTENCY and EVIDENCE one might think that the best option is to give up EVIDENCE. What the following case shows is that this will not get us out of trouble.

Belief, it is common to assume, is a relation that holds between an agent and an abstract object, viz., a proposition. Assuming this picture of belief we can show that CONSISTENCY and POSSIBILITY are inconsistent. Consider the following propositional analogue of β:

$(*)$ **Alpha doesn't believe the proposition expressed by** $(*)$[20]

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

Our transparency assumptions can be captured by the following analogues of (2) and (3):

(6) $B_\alpha \rho(*) \rightarrow B_\alpha \rho`(B_\alpha \rho(*))'$

(7) $\neg B_\alpha \rho(*) \rightarrow B_\alpha \rho`(\neg B_\alpha \rho(*))'$

[20] Again we should think of 'Alpha' as having a first-personal mode of presentation for Alpha.

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

(1f)	$\neg B_\alpha \rho(*)$	Assumption
(2f)	$B_\alpha \rho`\neg B_\alpha \rho(*)'$	(1f) and (B_{II})
(3f)	$B_\alpha \rho(*)$	(2f), (r1), substitution of equivalents
	$\bot$	

It follows that Alpha cannot fail to believe the proposition expressed by $(*)$. However, when Alpha believes this proposition, given (6), it is doomed to inconsistency. Thus:

(1g)	$B_\alpha \rho(*)$	Assumption
(2g)	$B_\alpha \rho`B_\alpha \rho(*)'$	(1g) and (B_I)
(3g)	$B_\alpha \rho`\neg B_\alpha \rho(*)'$	(1g), (r1), substitution of equivalents

Given CONSISTENCY, we will once again have a violation of POSSIBILITY. Holding fixed (6) and (7), 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 given the set of mutually exclusive and jointly exhaustive doxastic options consisting of believing the proposition expressed by $(*)$ and not believing this proposition, there is no option which it is possible for Alpha to realize and in so doing avoid incurring rational criticism. 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 out earlier case. The key difference is the replacement of '$T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$' 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 (1f)-(3f) and (1g)-(3g) involving the sentence β.

Corresponding to (1f)-(3f) we have:

(1h)	$\neg B_\alpha T(\beta)$	Assumption
(2h)	$B_\alpha \neg B_\alpha T(\beta)$	(1h) and (3)
(3h)	$B_\alpha T(\beta)$	(2h), (5), EVIDENCE
	$\bot$	

Corresponding to (1g)-(3g) we have:

(1i)	$B_\alpha T(\beta)$	Assumption
(2i)	$B_\alpha B_\alpha T(\beta)$	(1i) and (2)
(3i)	$B_\alpha \neg T(\beta)$	(2i), (4), EVIDENCE

Where in (1f)-(3f) appeal is made to the propositional identity $\rho(*) = \rho(\text{'}\neg B_\alpha \rho(*)\text{'})$, in (1h)-(3h) we must appeal to Alpha's justified belief in $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$, together with the assumption that Alpha meets the evidential norm EVIDENCE. The same is true of (1g)-(3g) and (1i)-(3i). 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.

§3 The Solution

I'll now show how certain resources that have been deployed to deal with the Liar paradox can be used to resolve the above two normative paradoxes. In particular, I'll show that Alpha can satisfy both CONSISTENCY and EVIDENCE, if we allow that it may be indeterminate whether Alpha believes that β is true. I'll focus primarily on the first paradox, noting later how the same treatment can be applied to the second.

It will help to first take a brief look at a related semantic paradox. Let 'η' name the following sentence: '$\neg\Box T(\eta)$'. On the assumption that the logic governing '$\Box$' is S5 we can derive a contradiction from this sentence as follows:

(1η)	$T(\eta) \leftrightarrow \neg\Box T(\eta)$	T-schema
(2η)	$\Box(\Box T(\eta) \to \neg T(\eta))$	(1η), Nec.
(3η)	$\Box(\neg\Box T(\eta) \to T(\eta))$	(1η), Nec.
(4η)	$\Box\Box T(\eta) \to \Box\neg T(\eta)$	(2η), K
(5η)	$\Box\neg\Box T(\eta) \to \Box T(\eta)$	(3η), K
(6η)	$\Box T(\eta) \to \Box\Box T(\eta)$	4
(7η)	$\neg\Box T(\eta) \to \Box\neg\Box T(\eta)$	5
(8η)	$\Box\neg T(\eta) \to \neg\Box T(\eta)$	S5 theorem
(9η)	$\Box T(\eta) \vee \neg\Box T(\eta)$	Classical Theorem

Subproof 1

(10η)	$\Box T(\eta)$	Assumption
(11η)	$\Box\Box T(\eta)$	(6η), (10η)
(12η)	$\Box\neg T(\eta)$	(4η), (11η)
(13η)	$\neg\Box T(\eta)$	(8η), (12η)
(14η)	$\bot$	(10η), (13η)

Subproof 2

(15η)	$\neg\Box T(\eta)$	Assumption
(16η)	$\Box\neg\Box T(\eta)$	(7η), (15η)
(17η)	$\Box T(\eta)$	(5η), (16η)
(18η)	$\bot$	(15η), (17η)
(19η)	$\bot$	(9η), (10η)-(14η), (15η)-(18η)

(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: $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$. (6η) holds if the logic governing '$\Box$' obeys axiom 4: $\Box\phi \to \Box\Box\phi$. (7η) holds if the logic governing '$\Box$' obeys axiom 5: $\neg\Box\phi \to \Box\neg\Box\phi$. (8η) holds given that axiom T: $\Box\phi \to \phi$ 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 $\Box T(\eta)$ and on the assumption $\neg\Box T(\eta)$ using simply *modus ponens*. A contradiction can then be derived outright from (9η) by proof-by-cases reasoning.

The approach to the Liar paradox that I'm interested in holds that excluded-middle is not valid for paradox inducing sentences. The above derivation is blocked at (9η). It is a fairly straightforward exercise to extend the class of models used to treat standard paradox inducing sentences such as the Liar sentence to models for languages containing '$\Box$'.[21] In any such paracomplete possible-worlds model in which accessibility between worlds is an equivalence relation, (1η)-(8η) all hold, but excluded-middle fails for '$\Box T(\eta)$'. Although we cannot say that η 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, i.e., it is indeterminate whether η is necessarily true. '$I\neg\Box T(\eta)$' is valid in the class of models.

The paradox developed in §2 can be represented in the form of a derivation that parallels the modal Liar paradox:

(1β)	$T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$	T-schema
(2β)	$B_\alpha(B_\alpha T(\beta) \rightarrow \neg T(\beta))$	Assumption
(3β)	$B_\alpha(\neg B_\alpha T(\beta) \rightarrow T(\beta))$	Assumption
(4β)	$B_\alpha B_\alpha T(\beta) \rightarrow B_\alpha \neg T(\beta)$	(2β), EVIDENCE
(5β)	$B_\alpha \neg B_\alpha T(\beta) \rightarrow B_\alpha T(\beta)$	(3β), EVIDENCE
(6β)	$B_\alpha T(\beta) \rightarrow B_\alpha B_\alpha T(\beta)$	Assumption
(7β)	$\neg B_\alpha T(\beta) \rightarrow B_\alpha \neg B_\alpha T(\beta)$	Assumption
(8β)	$B_\alpha \neg T(\beta) \rightarrow \neg B_\alpha T(\beta)$	CONSISTENCY
(9β)	$B_\alpha T(\beta) \vee \neg B_\alpha T(\beta)$	Classical Theorem

Subproof 1

(10β)	$B_\alpha T(\beta)$	Assumption
(11β)	$B_\alpha B_\alpha T(\beta)$	(6β), (10β)
(12β)	$B_\alpha \neg T(\beta)$	(4β), (11β)
(13β)	$\neg B_\alpha T(\beta)$	(8β), (12β)
(14β)	$\bot$	(10β), (13β)

[21]See the Technical Appendix for the formal details.

Subproof 2

(15β)	$\neg B_\alpha T(\beta)$	Assumption
(16β)	$B_\alpha \neg B_\alpha T(\beta)$	(7β), (15β)
(17β)	$B_\alpha T(\beta)$	(5β), (16β)
(18β)	$\perp$	(15β), (17β)
(19β)	$\perp$	(9β), (10β)-(14β), (15β)-(18β)

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

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 β is true then Alpha will have evidence that makes it certain that β is not true. So, assuming that Alpha meets the requirements imposed by EVIDENCE, it follows that if Alpha believes that it believes that β is true then it will believe that β is not true.[22] This is (4β). A similar story explains how (5β) follows from the assumption that Alpha meets the requirements imposed by EVIDENCE. (6β) and (7β) are assumptions that we made about Alpha. (8β) 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 $B_\alpha T(\beta)$ and on the assumption $\neg B_\alpha T(\beta)$. Given the assumption that the classical validity '$B_\alpha T(\beta) \vee \neg B_\alpha T(\beta)$' obtains, a contradiction can be derived outright by proof-by-cases reasoning.

Since the two derivations proceed in parallel, the strategy for blocking the contradiction in the former case will work equally well in the latter. Just as we can hold on to (1η)-(8η) by giving up excluded-middle for η, so too can we hold on to (1β)-(8β) by giving up excluded-

[22]This reasoning assumes that the deduction theorem holds. In fact, this is not so in the logic that I'll be adopting. Such reasoning is, however, "pre-theoretically valid" and so can be used in setting up a *prima facie* paradox. And, as we'll see the conditionals do in fact hold in the class of models in which I'm interested.

middle for β. If we allow that it may be indeterminate whether Alpha believes that β is true then we can block the derivation of a contradiction from the assumption that Alpha meets CONSISTENCY and EVIDENCE.

Postulating indeterminacy in Alpha's doxastic state allows us to defuse the argument given in §2.1 that CONSISTENCY, EVIDENCE and POSSIBILITY are incompatible. It is true that given that Alpha meets both CONSISTENCY and EVIDENCE we can derive a contradiction on the assumption that Alpha believes that β is true, and on the assumption that Alpha doesn't believe that β is true. This was taken to show that CONSISTENCY and EVIDENCE were in conflict with POSSIBILITY. Crucially, however, this requires that the pair of doxastic options consisting of believing that β is true and not believing that β is true are jointly exhaustive. But this is just the assumption that excluded-middle holds for the proposition that Alpha believes that β is true. If, then, excluded-middle fails for this proposition, it will not follow from the fact that we can derive a contradiction from each option, that there is a set of mutually exclusive and jointly exhaustive doxastic options no member of which can be realized without incurring rational criticism. We need not infer from the possibility of an agent such as Alpha that CONSISTENCY and EVIDENCE entail a violation of POSSIBILITY.

The paradox developed in §2.1 purported to show that there are possible cases in which CONSISTENCY and EVIDENCE cannot both be met. One way to think of the problem is as follows. Given that Alpha knows that $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$, it follows from the transparency assumptions, (2) and (3), that, whether or not Alpha believes that β 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 β is true) or leads to a violation of the requirement that the agent satisfy CONSISTENCY (in the case in which Alpha does believe that β is true).

Viewing the problem in terms of a local closure requirement lets us see more clearly how an appeal to indeterminacy can help resolve the problem. For, using the model theory developed to deal with the modal Liar sentence to interpret Alpha's belief operator, we can provide a model in which it is indeterminate whether Alpha believes that β is true and in which the following all hold[23]:

(8) $B_\alpha T(\beta) \leftrightarrow B_\alpha B_\alpha T(\beta)$

(9) $\neg B_\alpha T(\beta) \leftrightarrow B_\alpha \neg B_\alpha T(\beta)$

(10) $B_\alpha (T(\beta) \leftrightarrow \neg B_\alpha T(\beta))$

(11) $B_\alpha T(\beta) \rightarrow \neg B_\alpha \neg T(\beta)$

(12) Alpha's beliefs are closed under logical consequence

We can show, then, that if it is indeterminate whether or not Alpha believes that β is true, Alpha can satisfy the transparency assumptions, know that $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$, and yet not be forced into violating CONSISTENCY in order to meet EVIDENCE. This result does not depend on Alpha lacking transparency concerning its newly postulated indeterminate doxastic states. In the model in question the following also holds:

(13) $IB_\alpha T(\beta) \leftrightarrow B_\alpha IB_\alpha T(\beta)$

If, then, we allow that it is indeterminate whether Alpha believes that β is true we can hold on to what are seemingly quite plausible normative principles. And importantly for what will follow, it is also clear that for a proponent of a paracomplete treatment of the semantic paradoxes, this is *the only way* that one can hold on to CONSISTENCY, EVIDENCE and POSSIBILITY. For, on the assumption that Alpha meets CONSISTENCY and EVIDENCE, a paracomplete theorist will accept (1β)-(8β). 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 β.

Appeal to indeterminacy, then, allows us to hold on to CONSISTENCY and EVIDENCE in the face of the paradox developed in §2.1. Appeal to indeterminacy is also similarly effective in dealing

[23]See the Technical Appendix for proofs.

with the paradox developed in §2.2. There we derived a contradiction on the assumption that Alpha failed to believe the proposition expressed by (∗). We then inferred that Alpha must believe the proposition expressed by (∗). If, however, Alpha believes the proposition expressed by (∗), then Alpha will also believe its negation and so be in violation of CONSISTENCY. Alpha seemed doomed to irrationality, which by POSSIBILITY should not be possible.

The problem with this argument is to be located in the appeal to *reductio*. In the logics we are dealing with, *reductio* fails as a valid meta-rule. Where excluded-middle fails for ϕ, one cannot validly infer $\neg\phi$ given the derivation of a contradiction from ϕ. If, then, we allow that doxastic states may be subject to the same form of indeterminacy as cases of semantic paradox, it follows that while we can derive a contradiction from the assumption that Alpha fails to believe the proposition expressed by (∗) we cannot infer from this that Alpha does believe the proposition expressed by (∗). This suffices to block the conclusion that Alpha must be in violation of CONSISTENCY.

§4 Belief and Indeterminacy

I've argued that the normative paradoxes developed in §2 can be resolved if we allow that doxastic states may be indeterminate. But one may worry: Is it really possible for doxastic states to be indeterminate? The argument in this paper is addressed to one who accepts the possibility of indeterminacy at least with respect to cases of semantic pathology. So the worry is not germane if it stems from a general skepticism about the possibility of indeterminacy. The question, then, is whether one who accepts that semantic paradoxes give rise to indeterminacy should also allow that doxastic states may be indeterminate. In response to this question let me say two things.

(i) If one thinks that there is a univocal notion of indeterminacy that applies to cases of semantic pathology and to cases of vagueness, then there should be no worry at all about the possibility of doxastic indeterminacy. For vagueness is ubiquitous and there are certainly cases in which it is vague whether an agent has certain beliefs or not.

The question of whether there is a univocal notion of indeterminacy applicable to the semantic paradoxes and to cases of vagueness is subtle and it would take us too far afield at this point to make serious inroads on that question. But this unified view does seem to me to be an attractive option for one who supports an indeterminist treatment of the semantic paradoxes.[24]

(ii) Although the nature of doxastic states is an area of controversy, there are a number of attractive accounts of doxastic states that would make it intelligible, given what I've said so far, that such states could be subject to indeterminacy. A full development of any one of these accounts and how it is able to make sense of doxastic indeterminacy would require much more space than I have here. I can, however, briefly say something about the feature common to these views that let's us understand how doxastic states could be subject to indeterminacy.

The key feature of the accounts I have in mind is that they hold that principles of rationality are constitutive of intentional mental states such as belief and desire.[25] Given such an account, one should take doxastic states to be capable of having whatever properties are required by rationality. CONSISTENCY and EVIDENCE provide rational constraints on beliefs, and, as we've seen, in certain cases in order for an agent to conform to CONSISTENCY and EVIDENCE it must be indeterminate whether the agent believes a certain proposition. Rationality, thus, requires that an agent's doxastic states be in certain cases indeterminate. If one holds that principles of rationality are constitutive of doxastic states, then one should take such states to be, in principle, capable of exhibiting indeterminacy. Given such an account, it is a *discovery* afforded by the preceding arguments that doxastic states may exhibit the same form of indeterminacy that is present in cases of semantic pathology.[26]

Note that the thesis that doxastic states may be indeterminate poses no threat to physical-

[24]This is the position taken by Field. See, e.g., Field (2003b). Unified approaches to vagueness and semantic paradoxes are also developed in McGee (1991) and Soames (1999).

[25]One attractive view of this type has been developed by David Lewis. See Lewis (1974), Lewis (1999). According to Lewis belief and desire states are defined by a tacit theory of folk-psychology, and this tacit theory takes such states to conform to various rational principles. See also Stalnaker (1984) for a slightly different account which take conditions of rationality to be definitional of belief states. For another such account see Davidson (1980a,b).

[26]Note that I don't claim that theories of this type are the *only* theories that can admit doxastic indeterminacy. The claim is just that *given the preceding arguments* doxastic indeterminacy is intelligible for such theories.

ism. The account is perfectly compatible with the thesis that an agent's total doxastic state is identical to, say, a particular neural state of the agent. The point is just that for certain propositions, such as that expressed by β, that neural state will be such that it is indeterminate whether it is a state of believing the proposition in question.

There is much more to be said about the relationship between the nature of doxastic states and the postulation of doxastic indeterminacy. Our focus now, however, will be on the ways in which the postulation of doxastic indeterminacy in response to the normative paradox can provide insight into the correct answer to The Normative Question. I take it that this discussion is sufficiently motivated by the existence of at least two attractive (non-exclusive) views that countenance the existence of doxastic indeterminacy.

§5 Rejection

We've seen that if we allow that doxastic states may be indeterminate we can block the arguments given in §2 that purported to show the incompatibility of CONSISTENCY, EVIDENCE and POSSIBILITY. I'll now show how this approach to the normative paradoxes provides a constraint on the correct answer to The Normative Question. If we are to appeal to indeterminacy in order to resolve the normative paradoxes we must reject:

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.

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:

(14) $I\neg B_\alpha T(\beta) \leftrightarrow B_\alpha I\neg B_\alpha T(\beta)$

(14) holds in the class of 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.[27]

[27]See the Technical Appendix for the proof of this claim.

We have assumed that it is indeterminate whether Alpha does not believe that β is true, i.e., $I\neg B_\alpha T(\beta)$. By (14) it follows that Alpha believes this, i.e., $B_\alpha I\neg B_\alpha T(\beta)$. As in the earlier cases, we assume that Alpha is perfectly reliable in its beliefs about the indeterminacy of its own doxastic states. The following is a theorem: $I\neg B_\alpha T(\beta) \rightarrow IT(\beta)$.[28] 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 $IT(\beta)$. By EVIDENCE it follows that $OB_\alpha IT(\beta)$. By REJECTION, then, it follows that $OR_\alpha T(\beta)$. 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 $\neg B_\alpha T(\beta)$.

This, however, lands us back into normative paradox. We need simply rehearse Case 1. By (3) we have $B_\alpha \neg B_\alpha T(\beta)$. We also have that $\neg B_\alpha T(\beta) \rightarrow T(\beta)$ is a theorem, and that it is believed by Alpha on excellent grounds. It follows that Alpha's evidence makes it certain that $T(\beta)$. Assuming compliance with the normative demands imposed by EVIDENCE we have $B_\alpha T(\beta)$. But this, of course, is impossible since we already have $\neg B_\alpha T(\beta)$.

We have derived a contradiction on the assumption that Alpha meets both EVIDENCE, CONSISTENCY and REJECTION. Note that no appeal was made to excluded-middle.[29] 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 Normative Question. It is difficult to get an independent grip on this issue. Given this, we should, I think, take seriously

[28] In general where $\phi \leftrightarrow \psi$ is a theorem so is $I\phi \leftrightarrow I\psi$.

[29] Nor was any use made of *reductio* or other forms of proof which fail given the approach to the Liar under consideration.

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.

§6 Indeterminacy

If we reject REJECTION, what then is the correct answer to The Normative 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 β, another story is appropriate. Indeterminacy would in this way be like contingency.[30] 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 agnositic 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 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 propositions expressed by the Liar sentence is indeterminate, and the case in which it believes that the proposition expressed by β 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:

[30]Contingency is understood as follows: $C\phi \leftrightarrow_{df} \Diamond\phi \wedge \Diamond\neg\phi$.

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 INDETERMINACY in addition to CONSISTENCY and EVIDENCE.

§6.1

First, the argument for independent motivation.

It is very easy to be puzzled about what answer to give to The Normative Question. For, letting ϕ be some proposition that one ought to believe is indeterminate, *prima facie* the following three claims are all plausible consequences:

(15) One ought not believe ϕ.

(16) One ought not be agnostic about ϕ.

(17) One ought not reject, i.e., disbelieve, ϕ.

(15) 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 (16) maintains, to be inappropriate to be agnostic towards this proposition. Afterall, 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 this proposition about which they are ignorant.

(17) 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 $\neg\phi$.[31]

We are assuming that ϕ is indeterminate. If ϕ is indeterminate then so is $\neg\phi$. By (15), 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 is (17).

(15)-(17) would seem to exhaust the possible options. Obviously switching to a more fine-grained conception of doxastic states involving degrees of belief will not help, since such degrees will presumably be partitioned by the three states: belief, disbelief and agnosticism. *Prima facie*, then, it can seem that there is no rational attitude that one can take towards propositions that one should believe are indeterminate. Call this *The Normative Problem*.

An adequate response to the The Normative Problem should identify which of (15)-(17) we should give up, and *in addition* it should provide a plausible error-theory that can account for the *prima facie* plausibility of (15)-(17). What I will now do is outline such a response and show that INDETERMINACY is a consequence of this response. That INDETERMINACY follows from an elegant error-theoretic response to The Normative Problem gives us a reason to take INDETERMINACY seriously as the answer to The Normative Question.[32]

The response that I advocate involves rejecting each of (15)-(17). In their stead, we should accept the following closely related principles. Letting ϕ be a proposition that one ought to believe is indeterminate, I claim that the following all hold:

(15^d) One ought not *determinately* believe ϕ.

(16^d) One ought not be *determinately* agnostic about ϕ.

[31] Read: $O(R\phi \leftrightarrow B\neg\phi)$. This is a principle that a proponent of REJECTION will reject. But it should be conceded that this is *prima facie* quite plausible.

[32] I should note that this fact does not distinguish REJECTION and INDETERMINACY. For the proponent of REJECTION will reject NEGATION and can provide a plausible error-theory by (a) noting that NEGATION does hold in those cases in which indeterminacy is not present, and (b) appealing to the plausible fact that we are prone to overgeneralize from those cases in which indeterminacy isn't present. The point, then, is simply that the proponent of INDETERMINACY can also provide a well-motivated response to The Normative Problem.

(17^d) One ought not *determinately* reject ϕ.

Unlike with (15)-(17), an agent can meet each of the requirements imposed by (15^d)-(17^d). And a proponent of (15^d)-(17^d) can provide the following simple error-theory to account for the *prima facie* plausibility of (15)-(17). 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 (15^d)-(17^d), that concern certain conditions obtaining determinately, it should not be unexpected that we would confuse such principles for other principles, such as (15)-(17), that concern those conditions simply obtaining whether determinately or not. By accepting (15^d)-(17^d), then, we can account for the plausibility of (15)-(17), while avoiding its undesirable consequences.

Next I'll show that INDETERMINACY is a consequence of (15^d)-(17^d), i.e., that from (15^d)-(17^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 ($15^d$)-($17^d$).

(To 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 $OBI\phi \models OIB\phi$.)

Now, (i) just is (15^d), and so it trivially follows from (15^d)-(17^d).

To show that (ii) follows from (15^d)-(17^d), we'll need the assistance of the following principle:

SYMMETRY If one ought to believe that ϕ is indeterminate, then there should be no asymmetry between the determinate doxastic state one has concerning ϕ and the determinate doxastic state one has concerning $\neg\phi$.

SYMMETRY is, I think, quite a basic principle concerning how rational agents should respond to indeterminacy. For this reason I don't think that one can easily provide a justification for SYMMETRY by appeal to other clearer general principles. It is worth noting, however, that SYMMETRY is in fact a weakening of another *prima facie* plausible principle, one that is endorsed by a proponent of REJECTION. According to this stronger principle if one ought to believe that ϕ is indeterminate then there should be no asymmetry between the doxastic state one has concerning ϕ and the doxastic state that one has concerning $\neg\phi$. Anyone who endorses this latter principle should endorse SYMMETRY, although the converse does not hold.

Given SYMMETRY, we can show that (ii) follows from (15^d)-(17^d). First, note that it follows from (15^d)-(17^d) that: $OBI\phi \models O(\neg D\neg B\phi \vee \neg D\neg B\neg\phi)$.[33] Now, assume $OBI\phi$. It follows that $O(\neg D\neg B\phi \vee \neg D\neg B\neg\phi)$. Obviously, there are three ways of satisfying this latter obligation; one could satisfy one of the disjuncts but not the other, or one could satisfy both disjuncts. However, given that we have $OBI\phi$, it follows from SYMMETRY that the only appropriate way for our agent to satisfy this obligation is to satisfy both disjuncts. It follows that we have $O(\neg D\neg B\phi \wedge \neg D\neg B\neg\phi)$. Given ($15^d$)-($17^d$), $OBI\phi \models O(\neg D\neg B\phi \wedge \neg D\neg B\neg\phi)$. Since $O(\neg D\neg B\phi \wedge \neg D\neg B\neg\phi) \models O\neg D\neg B\phi$, we have $OBI\phi \models O\neg D\neg B\phi$. This suffices to establish (ii).

Since (i) and (ii) follow from (15^d)-(17^d), it follows that $OBI\phi \models OIB\phi$ is a consequence of (15^d)-(17^d). We've shown that INDETERMINACY follows from (15^d)-(17^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 Normative Question.

[33] Here's the argument. (16^d) tells us that $OBI\phi \models O\neg DA\phi$. To be agnostic about a proposition is to neither believe it nor reject it. We have, then, $OBI\phi \models O\neg D(\neg B\phi \wedge \neg R\phi)$. I'll take for granted NEGATION which tells us $O(B\neg\phi \leftrightarrow R\phi)$. Given that ought is closed under logical consequence we have the following general principle: $O(\psi \leftrightarrow \theta), O\gamma \models O\gamma_{\psi/\theta}$, where $\gamma_{\psi/\theta}$ is γ with one or more occurrences of ψ replaced by θ. It follows from this that if we have $\models O(\psi \leftrightarrow \theta)$ then we also have $O\gamma \models O\gamma_{\psi/\theta}$. We have, I am assuming, $\models O(B\neg\phi \leftrightarrow R\phi)$. Given this, then, it follows that $O\neg D(\neg B\phi \wedge \neg R\phi) \models O\neg D(\neg B\phi \wedge \neg B\neg\phi)$. From this and the fact that $OBI\phi \models O\neg D(\neg B\phi \wedge \neg R\phi)$, it follows that $OBI\phi \models O\neg D(\neg B\phi \wedge \neg B\neg\phi)$. In general, $D(\phi \wedge \psi)$ is equivalent to $D\phi \wedge D\psi$. In particular, then, $\neg D(\neg B\phi \wedge \neg B\neg\phi)$ is equivalent to $\neg(D\neg B\phi \wedge D\neg B\neg\phi)$ and so to $\neg D\neg B\phi \vee \neg D\neg B\neg\phi$. Since we can substitute logically equivalent sentences within the scope of 'O', we have, then, $OBI\phi \models O(\neg D\neg B\phi \vee \neg D\neg B\neg\phi)$.

§6.2

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 CONSISTENCY and EVIDENCE.

Alpha is an agent who believes the theorem $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$ and is, in addition, doxastically self-transparent. 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, given that it believes $T(\beta) \leftrightarrow \neg B_\alpha T(\beta)$ and is doxastically self-transparent. In such a model, however, it will be indeterminate whether Alpha believes that β 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 paracomplete possible-worlds models $\mathcal{M}$ the following holds[34]:

(18) $B_\alpha I\phi \models_{\mathcal{M}} IB_\alpha\phi$.

Given (18), 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.

Here's an argument for this claim.

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

[34]This assumes that B_α is treated in the models as a universal quantifier over possible-worlds. See the Technical Appendix for the proof of this claim

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, (18) 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 (18), 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.

§7 The Burden of Proof

I have argued, so far, that Agent Alpha can satisfy CONSISTENCY and EVIDENCE, given that its doxastic states are subject to indeterminacy. Alpha cannot, however, satisfy CONSISTENCY, EVIDENCE, and REJECTION. This, I've claimed, gives us good reason to reject REJECTION. Further, I've argued that if we allow that Alpha's doxastic states may be subject to indeterminacy, then it can satisfy CONSISTENCY, EVIDENCE and INDETERMINACY. The same problem that besets REJECTION does not beset INDETERMINACY. Given that INDETERMINACY is independently motivated, we have good reason to hold that it provides the correct answer to The Normative Question.

It has not been shown, however, that in *any* possible case in which an agent is not already guilty of a rational failure the agent may in principle satisfy CONSISTENCY, EVIDENCE and INDETERMINACY . Indeed, this is, I think, something that is not (at least in our present state of knowledge) amenable to proof. For both the conditions under which a proposition counts as part of an agent's evidence and the conditions under which an agent might count as being antecedently rational are in various ways unclear—the latter in particular depending on what other correct principles there are governing doxastic rationality. Trying, then, to prove in general that there are no cases in which an antecedently rational agent cannot meet CONSISTENCY, EVIDENCE and INDETERMINACY seems to me hopeless.

Given this limitation, it must be allowed that the arguments here are not unassailable. Perhaps there are cases in which an agent, not guilty of any antecedent rational failing, cannot satisfy CONSISTENCY and EVIDENCE despite the resources afforded by our paracomplete theory. Or perhaps the addition of INDETERMINACY may lead to cases in which an antecedently rational agent is condemned to irrationality. If the former were true, this would undermine our claim that we should reject REJECTION given the incompatibility of this principle with CONSISTENCY, EVIDENCE and POSSIBILITY. If the latter were true, this would undermine the claim that there is an important asymmetry between REJECTION and INDETERMINACY.

So far as I can see, however, we have no good reason to think that cases of either sort exist. And here it is worth emphasizing that despite the fact that we cannot prove *in general* that there are no such problematic cases, the arguments given above do generalize in important ways. In particular, the paracomplete model theory to which I have appealed is sufficient to assure us that there can be no case in which an agent is forced to violate CONSISTENCY, EVIDENCE and INDETERMINACY simply given knowledge of theorems and doxastic transparency. For, of course, in the models in which we have represented Alpha's doxastic state, such theorems will hold at every point. And this, we have seen, is compatible with the agent satisfying CONSISTENCY, EVIDENCE and INDETERMINACY. We can be assured then, not only that the case of Alpha provides no problem for the combination of CONSISTENCY, EVIDENCE and INDETERMINACY, but that, in general, there is no similar, but less obvious case, in which an agent is doomed to irrationality, by the lights of these principles, simply due to its knowledge of theorems and its own sensitivity to its doxastic states. This takes care of a large class of potentially problematic cases.

The burden of proof, at this point, seems to me to be squarely on the opponent of INDETERMINACY. Perhaps there are problematic cases of the sort described. However, until such a case is produced, we should, I think, invest a good amount of credence in INDETERMINACY.

§8 Rejection Again

I've argued that we have good reason to prefer INDETERMINACY to REJECTION. REJECTION, however, is not without its own positive motivations. In these final sections, I'll take up what are, I think, the three clearest arguments in favor of REJECTION and show how they can be resisted.

§8.1

Here is an argument which it might be thought strongly supports REJECTION.[35]

(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, e.g., $\lambda \wedge \neg\lambda$. In these cases, I hold that it should be indeterminate whether one rejects the contradiction in question.

[35] See Field (2003a) p.467 for this argument.

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 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.[36] 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).

[36] To 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 $R\phi$.

§8.2

Here's another argument in favor of REJECTION. This argument appeals to degrees of belief. So far I've confined myself to talking about binary belief. This has merely been for the sake of simplicity. Ultimately I think that binary belief should be understood in terms of degrees of belief. For the purposes of this argument let's make the following assumptions. To believe a proposition ϕ is to have a degree of belief above a certain threshold τ. To reject a proposition is to have a degree of belief below the co-threshold 1 - τ. 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:

(C2) If one has degree of belief over the threshold for $I\phi$, i.e., 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 say a bit about the form of the models which Field employs and how entailment is defined in these models. The models involve an infinite set of semantic values which are partially ordered.[37] There is a top value 1. Given

[37] These are the "fine-grained" semantic values. There is also another way one can describe the models in which there are only three values, with all the non-extremal values getting lumped together. To avoid confusion I note that, for reasons of simplicity, it is the corse-grained values that I employ in the Technical Appendix.

such a set of models we can define at least two notions of entailment. Call these respectively 'weak entailment' and 'strong entailment'.[38] We say that ϕ weakly entails ψ just in case in every model in which ϕ gets semantic value 1 ψ gets semantic value 1. We say that 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.[39] 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, i.e., 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, i.e., has semantic value 1 in every model. So, the only

[38]See Field (2008) p.169 for a discussion of these two notions of entailment.

[39]More 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.

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.

It is certainly true that we want a notion of entailment which 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 ϕ strongly entails ψ is equivalent to the claim that $\phi \rightarrow \psi$ is valid, i.e., has semantic value 1 in every model. The latter is equivalent to the claim that $T\ulcorner\phi\urcorner \rightarrow T\ulcorner\psi\urcorner$ 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, i.e., we cannot infer from $\phi \models \psi$ to $\models T\ulcorner\phi\urcorner \rightarrow T\ulcorner\psi\urcorner$. Since ϕ and $T\ulcorner\phi\urcorner$ 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.[40]

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.

§8.3

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 which 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.[41]

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 note first that an adequate treatment of the semantic paradoxes which appeals to indeterminacy requires that the indeterminacy operator iterate in a non-trivial man-

[40]See Field (2008) ch. 19 for an explanation of the relationship between the deduction theorem for weak validity and the Curry paradox.

[41]It is plausible to think first that objective chance cannot be analyzed in more basic terms and second that at 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 *exhausts* our understanding of objective chance. See Field (2003a) p. 479 for a comparison between indeterminacy and chance in this respect.

ner; we must not, e.g., have $II\phi \models I\phi$. In particular, this is required in order to adequately treat higher-order paradoxical sentences which employ the determinacy operator.

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 $II\phi \models I\phi$.

In general, for any value n, we can construct a higher-order paradoxical sentence which cannot be characterized as being indeterminate^n.[42] In order to characterize the distinctive paradoxical status of such sentences we require that it not be the case that $I^{n+1} \models I^n$.

An adequate treatment of the liar paradox which avails itself of the notion of indeterminacy requires that there be a hierarchy of non-equivalent indeterminacy operators. The question arises, then, what attitude should one have towards a proposition ϕ that one takes to be indeterminate^n for $n > 1$? A proponent of REJECTION should, I suggest, accept the following:

[42]In what follows I will restrict myself to consideration of cases in which n is a finite ordinal. Of course in certain cases we will need transfinite iterations of indeterminacy in order to characterize the status of certain paradoxical sentences. But the restriction will help simplify the argument and it is sufficient for the point I want to make.

REJECTION^n For every proposition ϕ and every $n \geq 1$ it is a consequence of the claim that one ought to believe that ϕ is indeterminate^n that one ought to reject ϕ and its negation.[43]

Here's an argument for this claim. 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. There is, however, a more general claim that it seems a proponent of REJECTION should accept, viz.:

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.[44]

Given this more general rejectionist claim REJECTION^n can be shown to hold. To show that REJECTION^n holds it suffices to establish:

(19) $OBI^{n+1}\phi \models OR\neg I^n\phi$

(20) $OR\neg I^n \models OR\phi \wedge OR\neg\phi$

(19) is an obvious consequence of REJECTION. The latter tells us that $OBI^{n+1}\phi \models ORI^n\phi \wedge OR\neg I^n\phi$. Since $ORI^n\phi \wedge OR\neg I^n\phi \models OR\neg I^n\phi$, by transitivity of entailment we have (19).

(20) can be established by a simple inductive argument using the following facts.

(21) $OR\neg I\phi \models OR\phi \wedge OR\neg\phi$[45]

(22) $OR\neg I^{n+1}\phi \models OR\neg I^n\phi$[46]

[43]Read: $OBI^n\phi \models OR\phi \wedge OR\neg\phi$.

[44]Read: $ORD\phi \wedge ORD\neg\phi \models OR\phi \wedge OR\neg\phi$. Field, e.g., would accept this more general statement. For, as noted, he thinks that it is a rational requirement that one have the same degree of belief in ϕ and its negation.

[45]To see that (21) holds 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 (21).

[46]To see that (22) holds note that from the equivalence of $\neg I\phi$ and $D\phi \vee D\neg\phi$ it follows that $OR\neg I^{n+1} \models OR(DI^n\phi \vee D\neg I^n\phi)$. Given $OR(\gamma \vee \psi) \models OR\gamma \wedge OR\psi$, we have $OR(DI^n\phi \vee D\neg I^n\phi) \models ORDI^n\phi \wedge ORD\neg I^n\phi$. And so we have $OR\neg I^{n+1} \models ORDI^n\phi \wedge ORD\neg I^n\phi$. By REJECTION′ we have $ORDI^n\phi \wedge ORD\neg I^n\phi \models ORI^n\phi \wedge OR\neg I^n\phi$. And so we have $OR\neg I^{n+1} \models ORI^n\phi \wedge OR\neg I^n\phi$ and thus $OR\neg I^{n+1}\phi \models OR\neg I^n\phi$, which is (22).

If $n = 1$ then (20) follows from (21). Next assume, as an induction hypothesis, that (20) holds for n. By (22) it follows that (20) holds for $n + 1$. So (20) holds in general.

We can now argue for REJECTION^n as follows. Obviously if $n = 1$ then then we have $OBI^n\phi \models OR\phi \wedge OR\neg\phi$ by REJECTION. So let $n = m+1$ for some positive m. In this case $OBI^n\phi \models OR\phi \wedge OR\neg\phi$ follows from (19) and (20). So $OBI^n\phi \models OR\phi \wedge OR\neg\phi$ holds in general for arbitrary ϕ and $n \geq 1$, which suffices to establish REJECTION^n.

We are now in a position to see why the explanatory asymmetry between the proponent of REJECTION and INDETERMINACY is not as great as it might first appear. It is true that REJECTION gives us a grip on indeterminacy in terms which do not presuppose an understanding of indeterminacy. Nonetheless REJECTION does not distinguish between indeterminacy and indeterminate indeterminacy or between indeterminacy and indeterminate indeterminate indeterminacy etc., for there are equivalent principles governing these operators. 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.

How then do we understand the difference? This is, I think, a very difficult question. But what I think a proponent of REJECTION should say, at least in outline, is that we are able to grasp the difference between these by understanding (a) the different paradigm cases in which they apply and (b) the difference in the logical behavior of the two operators. So while a proponent of REJECTION can say that we get some grip on the concept of indeterminacy in virtue of its cognitive role this is only a partial grip. What completes our understanding is our grasp of the paradigm cases in which indeterminacy applies together with our understanding of the basic logical principles governing this operator.

The difference, then, between the proponent of REJECTION and myself is that I must hold that our understanding of indeterminacy comes solely from the latter two sources. 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 the concept. 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 which point strongly in favor of INDETERMINACY then this is a cost, I think, worth incurring.

§9 Conclusion

We started with The Normative Question: What attitude should a rational agent take towards a proposition that it takes to be indeterminate? The answer to this question is, I claimed, not at all obvious. Nonetheless, there has been a strong consensus that the correct answer is provided by REJECTION. I've argued, however, that attention to the normative paradox raised by Agent Alpha motivates, instead, INDETERMINACY as the answer to this question.

The case of Agent Alpha showed that three plausible principles of rationality, CONSISTENCY, EVIDENCE and POSSIBILITY, were at least *prima facie* inconsistent. This particular paradox, however, can be resolved if we allow that doxastic states are subject to indeterminacy. If it is indeterminate whether Alpha believes that β is true, then Alpha will be able to satisfy both CONSISTENCY and EVIDENCE. This solution, however, is undermined if one accepts REJECTION. Since CONSISTENCY, EVIDENCE and POSSIBILITY are all much more intuitively plausible than REJECTION, this provides a strong *prima facie* reason to reject REJECTION. The same problem, however, does not beset INDETERMINACY. Moreover, INDETERMINACY can be given independent motivation, since it provides an attractive error-theoretic response to The Normative Problem. All of this provides, I claim, a compelling cumulative case to think that the orthodox answer to The Normative Question is mistaken. We should instead hold that the correct answer is given by INDETERMINACY.

According to the picture that emerges, indeterminacy in the objects of a rational agent's doxastic states will filter up to the attitudes themselves. Such an agent's doxastic states will exhibit the same paradoxical features as the objects of its attitudes. If this is right, then the study of the semantic paradoxes has much broader implications then has traditionally been thought. Attention to the semantic paradoxes can provide surprising insights into the nature of rationality and of the mental states of rational agents.

§A Technical Appendix

In this appendix, I outline the underlying technical machinery appealed to in the body of the paper and prove certain results that are important for the arguments therein.

§A.1 General Framework

The model theory that I'll sketch uses tools developed by Hartry Field and extends them to treat languages involving modal operators. For a full development of Field's model theory see chapters 15-17 in Field (2008).

The Field-style models that we'll be constructing make essential use of the fixed-point construction developed in Kripke (1975). Let me first give a brief sketch of how this construction works for a language involving modal operators. For a detailed exposition of how such constructions proceed (for languages not involving modal operators) see Kripke (1975).[47]

Kripke Models: Let us start with a standard classical model M for a language L containing a modal operator '$\Box$', but not containing a truth predicate. M will be a quadruple $< D_m, \Delta_m, R_m, I_m >$. D_m is the domain of individuals. Δ_m is a set of points relative to which sentences are assigned truth-values. R_m is a relation of "accessibility" holding between members of Δ. I_m is an interpretation function that assigns classical values to the elements of M relative to members of Δ. Truth relative to a point and a sequence is defined in the standard way.

Let L^+ be the language that results by adding a truth predicate 'T' to L. I'll now show how we can extend M to a model M^+ for the language L^+. M^+ will be a non-classical model. M^+ will be identical to M except that I_m^+ will assign a semantic value to 'T' relative to members

[47] There are in addition a number of secondary source expositions. Two useful sources are chapter 3 in Field (2008), or chapter 6 in Soames (1999).

of Δ. Unlike in a classical model, however, 'T' is not assigned a single extension relative to a point δ. Instead it is assigned an extension $T^{\delta+}$ and an anti-extension $T^{\delta-}$. $T^{\delta+}$ will be a set of sentences. $T^{\delta-}$ will be a set consisting of all non-sentences together with any sentence whose negation is in $T^{\delta+}$. Not every sentence will be in $T^{\delta+} \cup T^{\delta-}$. In this model, sentences can receive one of three semantic values relative to a point of evaluation: 1, $1/2$, or 0. Sentences in the base language L will receive only values 1 and 0, but sentences involving the new predicate T may receive value $1/2$.

M^+ is constructed by considering a series of models M^+_α. Each M^+_α is, like M^+, identical to M except that I^+_α assigns an extension and anti-extension to 'T' relative to each member of Δ. We let $T^{\delta+}_\alpha$ denote the extension of 'T' under M^+_α. $T^{\delta-}_\alpha$ will denote the anti-extension as defined above.

We first provide an inductive definition of what it is for a formula γ of L^+ to have a semantic value 1, $1/2$, or 0 relative to a sequence s and point δ under M^+_α.

- If γ is a formula of the form $P(t_1, t_2.....t_n)$, where P is a predicate of L, then:

  $[\![P(t_1, t_2.....t_n)]\!]^{s,\delta,M^+_\alpha} = 1$ iff $< [\![t_1]\!]^{s,\delta,M^+_\alpha}, ..., [\![t_n]\!]^{s,\delta,M^+_\alpha} > \in [\![P]\!]^{s,\delta,M^+_\alpha}$

  Otherwise $[\![P(t_1, t_2.....t_n)]\!]^{s,\delta,M^+_\alpha} = 0$

- If γ is a formula of the form $T(t)$ then:

  $[\![T(t)]\!]^{s,\delta,M^+_\alpha} = 1$ iff $[\![t]\!]^{s,\delta,M^+_\alpha} \in T^{\delta+}_\alpha$

  $[\![T(t)]\!]^{s,\delta,M^+_\alpha} = 0$ iff $[\![t]\!]^{s,\delta,M^+_\alpha} \in T^{\delta-}_\alpha$

  Otherwise $[\![T(t)]\!]^{s,\delta,M^+_\alpha} = 1/2$

- If γ is a formula of the from $\neg\phi$ then:

  $[\![\neg\phi]\!]^{s,\delta,M^+_\alpha}$ is equal to $1 - [\![\phi]\!]^{s,\delta,M^+_\alpha}$

- If γ is a formula of the from $\phi \wedge \psi$ then:

  $[\![\phi \wedge \psi]\!]^{s,\delta,M^+_\alpha}$ is equal to $\min\{[\![\phi]\!]^{s,\delta,M^+_\alpha}, [\![\psi]\!]^{s,\delta,M^+_\alpha}\}$

- If γ is a formula of the from $\phi \vee \psi$ then:

  $[\![\phi \vee \psi]\!]^{s,\delta,M_\alpha^+}$ is equal to $\max\{[\![\phi]\!]^{s,\delta,M_\alpha^+}, [\![\psi]\!]^{s,\delta,M_\alpha^+}\}$

- If γ is a formula of the from $\forall x\phi$ then:

  $[\![\forall x\phi]\!]^{s,\delta,M_\alpha^+}$ is equal to $\min\{[\![\phi]\!]^{s',\delta,M_\alpha^+} : s' \in \Sigma^{s/x}\}$, where $\Sigma^{s/x}$ is the set of sequences differing from s at most with respect to x.

- If γ is a formula of the form $\Box\phi$ then:

  $[\![\Box\phi]\!]^{s,\delta,M_\alpha^+}$ is equal to $\min\{[\![\phi]\!]^{s,\delta',M_\alpha^+} : \delta' \in \Delta^\delta\}$, where Δ^δ is the set of points δ' such that $\delta R_m \delta'$.

We say that a sentence ϕ has a semantic value x relative to a model M_α^+ and a point δ, $[\![\phi]\!]^{\delta,M_\alpha^+} = x$, iff for all s $[\![\phi]\!]^{s,\delta,M_\alpha^+} = x$. Note that if ϕ is a sentence then $[\![\phi]\!]^{s,\delta,M_\alpha^+} = [\![\phi]\!]^{s',\delta,M_\alpha^+}$ for all sequences s and s'.

We now construct a series of models M_α^+. We start by setting $T_0^{\delta+} = \emptyset$ for all δ. At stage $\alpha + 1$ we let $T_{\alpha+1}^{\delta+}$ be the set of sentences ϕ such that $[\![\phi]\!]^{\delta,M_\alpha^+} = 1$. At a limit stage λ we let $T_\lambda^{\delta+}$ be the set of sentences ϕ such that for some $\beta < \lambda$ $[\![\phi]\!]^{\delta,M_\beta^+} = 1$.

FIXED POINT THEOREM There exists a least ordinal σ such that for every δ the set of sentences ϕ such that $[\![\phi]\!]^{\delta,M_\sigma^+} = 1$ is identical to $T_\sigma^{\delta+}$.

We let M^+ be M_σ^+.

Proof Sketch for the FIXED POINT THEOREM: Take some arbitrary well-ordering of the set Δ_m. This will associate with each $\delta \in \Delta_m$ a corresponding ordinal up to some ordinal ξ.[48] Using this well-ordering we can associate with any model M_α^+ a sequence $< A_\beta : \beta \in \xi >^{M_\alpha^+}$, where A_β is the set of sentences which M_α^+ assigns as an extension for 'T' relative to the point associated with ordinal β.

[48]By Zermelo's Theorem any set can be well-ordered. Δ_m will be set-sized. It follows that there will be some well-ordering of this set.

Let $\leq$ be the following partial-order on such sequences: $< A_\beta : \beta \in \xi > \leq < B_\beta : \beta \in \xi >$ iff for every $\beta \in \xi$ $A_\beta \subseteq B_\beta$.

A simple inductive argument will show that the following property holds for every formula ϕ:

(**) For every point δ and sequence s, if $< A_\beta : \beta \in \xi >^{M^+_\alpha} \leq < B_\beta : \beta \in \xi >^{M^+_\beta}$ then (a) if $[\![\phi]\!]^{s,\delta,M^+_\alpha} = 1$ then $[\![\phi]\!]^{s,\delta,M^+_\beta} = 1$ and (b) if $[\![\phi]\!]^{s,\delta,M^+_\alpha} = 0$ then $[\![\phi]\!]^{s,\delta,M^+_\beta} = 0$.[49]

It is a consequence of (**) that in the above construction if $\alpha \leq \beta$ then $< A_\beta : \beta \in \xi >^{M^+_\alpha}$ $\leq < B_\beta : \beta \in \xi >^{M^+_\beta}$. As our construction proceeds the extensions assigned to 'T' relative to a point will either stay the same or will grow. Growth, however, cannot continue indefinitely. Here's a way of seeing why this is the case. Imagine a construction in which we added at each stage a single sentence to the extension of 'T' at a single point—the minimal amount of growth possible. Given such a construction there will be some ordinal by which we will have run out sentences to add. Let C^S be the cardinality of the set of sentences. We can think of ourselves as having a number of such sets, viz., one for every point in Δ_m. Let C^{Δ_m} be the cardinality of Δ_m. In total we have a set of sentences of cardinality $C^S \times C^{\Delta_m}$. There are, however, ordinals of greater cardinality than this. Once we have reached an ordinal of size greater than $C^S \times C^{\Delta_m}$ we will have run out of sentences to add to points. There will be some ordinal σ, then, such that $< A_\beta : \beta \in \xi >^{M^+_\sigma} = < B_\beta : \beta \in \xi >^{M^+_{\sigma+1}}$. And since the ordinals are well-ordered there will be a least such ordinal.

Note that under M^+ a sentence ϕ and $T\ulcorner\phi\urcorner$ will have the same semantic value relative to a point δ and a sequence s, where $\ulcorner\phi\urcorner$ is any term that denotes ϕ under M^+ relative to δ and s. In general then substituting an occurrence of ϕ for $T\ulcorner\phi\urcorner$ in a sentence ψ will not change the

[49]For the sake of concision I omit the proof of this here. The proof, however, is a fairly straightforward generalization of proof, given in Kripke (1975), of the monotonicity of the function mapping extensions of 'T' to formulas having semantic value 1 given that assignment.

semantic value of ψ under M^{+} relative to δ and s.

Field Models: We will now show how to construct a model for a language L^{++}, which, like L^{+}, contains a truth predicate, but in addition contains a conditional '$\rightarrow$'. (Using the resources of this language, a determinacy operator 'D' can be defined as $D\phi \leftrightarrow_{df} \phi \wedge \neg(\phi \rightarrow \neg\phi)$.) The construction developed by Field involves a transfinite sequence of Kripke constructions for a language not involving a modal operator. It is straightforward to apply the same sort of construction using instead a transfinite sequence of Kripke constructions for a language involving a modal operator.

The construction works as follows. At each stage in the sequence we begin with a model M^{α}. M^{α} assigns to the elements of language L the assignments provided by M, it assigns to 'T' the nullset as extension at each point δ and, in addition, it assigns, relative to a sequences s and point δ, semantic values to formulas which have '$\rightarrow$' as their main connective. Given such a starting model we then construct a Kripke model M^{α}_{+} using the method described above.[50]

Consider an arbitrary formula with '$\rightarrow$' as its main connective: $\phi \rightarrow \psi$. The assignment $[\![\phi \rightarrow \psi]\!]^{s,\delta,M^{\alpha}}$ is determinate as follows:

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

[50]Typography note. Do not confuse M^{α}_{+} with M^{+}_{α}. The latter is the model at the α stage of the Kripke construction. The former, which we are now considering, is the Kripke model at the α stage of Field's construction.

At each stage α, then, a formula ϕ will receive a semantic value relative to a sequence s and point δ, given the resulting Kripke model at that stage M_{+}^{α}. Certain formulas will at some point in this series stabilize at value 1 (relative to s and δ). We say that such formulas have ultimate value 1 (relative to s and δ). Other formulas will eventually stabilize at value 0 (relative to s and δ). In this case we say that such formulas have ultimate value 0 (relative to s and δ). Other formulas will either eventually stabilize at value $1/2$ (relative to s and δ) or will never stabilize at any value. These formulas we say have ultimate value $1/2$ (relative to s and δ). Following Field we denote the ultimate value of a formula relative to a sequence s and point δ $|||\phi|||^{s,\delta}$.

This resulting assignment of values gives us our model. Ultimate value 1 is the designated value. We can then define validity within a class of modal Field-models $\mathcal{M}$ as follows. We say that for sentences ϕ and ψ, $\phi \models_{\mathcal{M}} \psi$ iff for every $M \in \mathcal{M}$, and every $\delta \in \Delta_m$, if $|||\phi|||^{\delta,M} = 1$ then $|||\psi|||^{\delta,M} = 1$.

I note that the following result holds:

FUNDAMENTAL THEOREM For any ordinal ρ there exists an ordinal $\xi > \rho$ such that for every formula ϕ, sequence s and point δ, if $|||\phi|||^{s,\delta} = x$ then $[\![\phi]\!]^{s,\delta,M_{+}^{\xi}} = x$.[51]

Ordinals such as ξ are called 'acceptable'. The existence of such ordinals ensures that, like the Kripke models, the logic induced by this class of models for the fragment L^{+} is K_3.

§A.2 Some Theorems

Having outlined the manner in which our models are constructed, I will now justify certain claims made in the paper about these models.

First I'll justify claims (8)-(14).

Let M be a classical possible worlds model for which R_m is an equivalence relation, i.e., a transitive, reflexive and symmetric relation. Let B_α be treated as $\Box$. For all s and δ the following hold in the model generated by the Field construction.

[51] Again for the sake of concision this proof is omitted here. The proof is a fairly straightforward generalization of the proof of the Fundamental Theorem given in Field (2008).

(8) $|||B_\alpha T(\beta) \leftrightarrow B_\alpha B_\alpha T(\beta)|||^{s,\delta} = 1$.

For each stage α in the construction $[\![B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = [\![B_\alpha B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$. So at each stage $\alpha > 0$ $[\![B_\alpha T(\beta) \leftrightarrow B_\alpha B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = 1$.

(9) $|||\neg B_\alpha T(\beta) \leftrightarrow B_\alpha \neg B_\alpha T(\beta)|||^{s,\delta} = 1$.

As above, for each stage α in the construction $[\![\neg B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} =$

$[\![B_\alpha \neg B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$. So at each stage $\alpha > 0$ $[\![\neg B_\alpha T(\beta) \leftrightarrow B_\alpha \neg B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$

$= 1$.

(10) $|||B_\alpha(T(\beta) \leftrightarrow \neg B_\alpha T(\beta))|||^{s,\delta} = 1$.

For each stage $\alpha > 0$ in the construction and each point δ', $[\![T(\beta) \leftrightarrow \neg B_\alpha T(\beta)]\!]^{s,\delta',M_+^\alpha} = 1$. It follows that for every stage $\alpha > 0$ $[\![B_\alpha(T(\beta) \leftrightarrow \neg B_\alpha T(\beta))]\!]^{s,\delta,M_+^\alpha} = 1$.

(11) $|||B_\alpha T(\beta) \rightarrow \neg B_\alpha \neg T(\beta)|||^{s,\delta} = 1$.

Assume $[\![B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = 1$. It follows that for every point δ' $[\![T(\beta)]\!]^{s,\delta',M_+^\alpha}$

$= 1$, and so $[\![\neg T(\beta)]\!]^{s,\delta',M_+^\alpha} = 0$. It follows that $[\![B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha} = 0$ and so $[\![\neg B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha} =$

1.

Assume $[\![B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = 1/2$. It follows that for no point δ' is it the case that $[\![T(\beta)]\!]^{s,\delta',M_+^\alpha} =$

0 and for some point δ' it is the case that $[\![T(\beta)]\!]^{s,\delta',M_+^\alpha} = 1/2$. From this it follows that every point δ' is such that either $[\![\neg T(\beta)]\!]^{s,\delta',M_+^\alpha}$

$= 0$ or $[\![\neg T(\beta)]\!]^{s,\delta',M_+^\alpha} = 1/2$. Thus, $[\![B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha} \leq 1/2$, and so $[\![\neg B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha} \geq$

$1/2$.

For every stage α, then, $[\![B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} \leq [\![\neg B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha}$. And so for every stage $\alpha > 0$ $[\![B_\alpha T(\beta) \rightarrow \neg B_\alpha \neg T(\beta)]\!]^{s,\delta,M_+^\alpha} = 1$.

(12) For any $\mathcal{M}$, $\phi \models_{\mathcal{M}} \psi \Rightarrow B_\alpha \phi \models_{\mathcal{M}} B_\alpha \psi$.

Assume that $\phi \models_{\mathcal{M}} \psi$. Take an arbirary $M \in \mathcal{M}$ and assume that in this model $|||B_\alpha \phi|||^\delta =$

1. We'll argue that $|||B_\alpha \psi|||^\delta = 1$. At every point δ' such that $\delta R \delta'$, $|||\phi|||^\delta = 1$. Given that $\phi \models_{\mathcal{M}} \psi$ it follows that at every such δ' $|||\psi|||^{\delta'} = 1$. And so $|||B_\alpha\psi|||^{\delta} = 1$. Thus $B_\alpha\phi \models_{\mathcal{M}} B_\alpha\psi$.

(13) $|||IB_\alpha T(\beta) \leftrightarrow B_\alpha IB_\alpha T(\beta)|||^{s,\delta} = 1$.

Let α be an arbitrary ordinal.

I claim that $[\![IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = [\![B_\alpha IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$. To see this note that $IB_\alpha T(\beta)$ has the same semantic value relative to each point in Δ_m, given M_+^α . For we have $IB_\alpha T(\beta) \leftrightarrow_{df} \neg DB_\alpha T(\beta) \wedge \neg D\neg B_\alpha T(\beta) \leftrightarrow_{df} \neg(B_\alpha T(\beta) \wedge \neg(B_\alpha T(\beta) \rightarrow \neg B_\alpha T(\beta))) \wedge \neg(\neg B_\alpha T(\beta) \wedge \neg(\neg B_\alpha T(\beta) \rightarrow B_\alpha T(\beta)))$. Given that R_m is an equivalence relation it follows that each of $B_\alpha T(\beta)$, $B_\alpha T(\beta) \rightarrow \neg B_\alpha T(\beta)$, and $\neg B_\alpha T(\beta) \rightarrow B_\alpha T(\beta)$ have the same semantic value relative to each point of evaluation, given M_+^α. This is sufficient to guarantee that $IB_\alpha T(\beta)$ has the same semantic value x relative to each point of evaluation. And whatever value x is that will be the same value that $B_\alpha IB_\alpha T(\beta)$ has at every point of evaluation. In particular, then, $[\![IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = [\![B_\alpha IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$. It follows that for every $\alpha > 0$

$[\![IB_\alpha T(\beta) \leftrightarrow B_\alpha IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} = 1$.

(14) $|||I\neg B_\alpha T(\beta) \leftrightarrow B_\alpha I\neg B_\alpha T(\beta)|||^{s,\delta} = 1$

The proof of this is essentially the same as the proof of (13). It suffices to note that for any model M_+^α and any point δ, $[\![IB_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha} =$

$[\![I\neg B_\alpha T(\beta)]\!]^{s,\delta,M_+^\alpha}$.

Finally, I provide a justification for claim (18).

(18) For any class of modal Field-models $\mathcal{M}$, $B_\alpha I\phi \models_{\mathcal{M}} IB_\alpha\phi$.

Let M be some model in $\mathcal{M}$. Assume that in this model $|||B_\alpha I\phi|||^{\delta} = 1$. We'll now show that given this assumption $|||IB_\alpha\phi|||^{\delta} = 1$

By the FUNDAMENTAL THEOREM we know that there is a class of acceptable ordinals. Let's denote the least such ordinal 'σ'. To show that $|||IB_\alpha\phi|||^{\delta} = 1$ follows on the assumption that $|||B_\alpha I\phi|||^{\delta} = 1$ it will suffice to establish:

(D1) If $|||B_\alpha I\phi|||^{\delta} = 1$ then for all δ' such that $\delta R_m \delta'$, for all $\xi > \sigma$, $[\![\phi]\!]^{\delta', M^{\xi}_{+}} = 1/2$.

For if for all δ' such that $\delta R_m \delta'$, for all $\xi > \sigma$, $[\![\phi]\!]^{\delta', M^{\xi}_{+}} = 1/2$, then it follows that for all $\xi > \sigma$, $[\![B_\alpha \phi]\!]^{\delta, M^{\xi}_{+}} = 1/2$. And from this it follows that for all $\xi > \sigma + 2$, $[\![IB_\alpha \phi]\!]^{\delta, M^{\xi}_{+}} = 1$. This can be verified by noting that for any model M^{ξ}_{+} and point δ', $I\phi$ will receive the same semantic value as $(\neg\phi \vee \phi \rightarrow \neg\phi) \wedge (\phi \vee \neg\phi \rightarrow \phi)$.

To show that (D1) holds we will establish:

(D2) If $|||B_\alpha I\phi|||^{\delta} = 1$ then for all δ' such that $\delta R_m \delta'$ $|||I\phi|||^{\delta'} = 1$.

(D3) If $|||I\phi|||^{\delta'} = 1$, then for all $\xi > \sigma$, $[\![\phi]\!]^{\delta', M^{\xi}_{+}} = 1/2$.

Justification for (D2): (D2) is a straightforward consequence of our treating B_α as a universal quantifier over the set of δ' such that $\delta R_m \delta'$.

Justification for (D3): Assume $|||I\phi|||^{\delta'} = 1$. We'll show that on this assumption it follows that for all $\xi > \sigma$, $[\![\phi]\!]^{\delta', M^{\xi}_{+}} = 1/2$. We do so by reductio. We'll assume that there is some $\xi > \sigma$ $[\![\phi]\!]^{\delta', M^{\xi}_{+}} \neq 1/2$ and show that from this, together with our other assumption, a contradiction can be derived. This is sufficient to establish (D3).

To argue for this we first establish the following:

(D4) If $|||I\phi|||^{\delta'} = 1$, then for all $\xi > \sigma$, if $\exists\sigma'\forall\sigma''$ $\sigma' \leq \sigma'' < \xi$ $[\![\phi]\!]^{\delta', M^{\sigma''}_{+}} \neq 1/2$, then $[\![\phi]\!]^{\delta', M^{\xi}_{+}} \neq 1/2$.

Justification for (D4): Assume that $|||I\phi|||^{\delta'} = 1$ and that $\exists\sigma'\forall\sigma''$ $\sigma' \leq \sigma'' < \xi$ $[\![\phi]\!]^{\delta', M^{\sigma''}_{+}} \neq 1/2$. If $|||I\phi|||^{\delta'} = 1$, then for all $\xi > \sigma$ $[\![I\phi]\!]^{\delta', M^{\xi}_{+}} = 1$.[52] Our first assumption, then, guarantees that $[\![I\phi]\!]^{\delta', M^{\xi}_{+}} = 1$ Given this and our second assumption it follows that $[\![\phi]\!]^{\delta', M^{\xi}_{+}} \neq 1/2$. To see this note that for any model M^{ξ}_{+} and point δ', $I\phi$ will receive the same semantic value as $(\neg\phi \vee \phi \rightarrow \neg\phi) \wedge (\phi \vee \neg\phi \rightarrow \phi)$. So we have $[\![(\neg\phi \vee \phi \rightarrow$

[52] To see this note that the sequence of Kripke constructions 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 reoccur and this will institute a cyclical pattern. And such a pattern will have been instituted by the time the first acceptable ordinal occurs, since this is one of the reoccurring valuations.

$\neg\phi) \wedge (\phi \vee \neg\phi \rightarrow \phi)]\!]^{\delta',M^{\xi}_{+}} = 1$. But it is clear that given that $\exists\sigma'\forall\sigma''\ \sigma' \leq \sigma'' < \xi$ $[\![\phi]\!]^{\delta',M^{\sigma''}_{+}} \neq 1/2$, then in order for the above formula to have value 1, either $[\![\phi]\!]^{\delta',M^{\xi}_{+}} = 1$ or $[\![\phi]\!]^{\delta',M^{\xi}_{+}} = 0$. For given that $\exists\sigma'\forall\sigma''\ \sigma' \leq \sigma'' < \xi\ [\![\phi]\!]^{\delta',M^{\sigma''}_{+}} \neq 1$, we cannot have both $[\![\phi \rightarrow \neg\phi]\!]^{\delta',M^{\xi}_{+}} = 1$ and $[\![\neg\phi \rightarrow \phi]\!]^{\delta',M^{\xi}_{+}} = 1$. In order, then, for both disjuncts to have value 1, ϕ must either have value 1 or value 0. This suffices to establish (D4).

From (D4), together with our assumption that $|||I\phi|||^{\delta'} = 1$, and our assumption that there is some $\xi > \sigma\ [\![\phi]\!]^{\delta',M^{\xi}_{+}} \neq 1/2$, it follows that for all $\xi' \geq \xi\ [\![\phi]\!]^{\delta',M^{\xi'}_{+}} \neq 1/2$. By the FUNDAMENTAL THEOREM there exists an acceptable ordinal σ' such that $\sigma' > \sigma$. We have that either $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$ or $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 0$. Both of these, however, conflict with our assumption that $|||I\phi|||^{\delta'} = 1$. Given that $|||I\phi|||^{\delta'} = 1$ it follows that $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$, and so $[\![(\neg\phi \vee \phi \rightarrow \neg\phi) \wedge (\phi \vee \neg\phi \rightarrow \phi)]\!]^{\delta',M^{\sigma'}_{+}} = 1$. But this cannot be the case if either $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$ or $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 0$. Assume $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$. Then in order for $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$ it must be that $[\![\phi \rightarrow \neg\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$. However, since σ' is an acceptable ordinal it follows that $|||\phi|||^{\delta'} = 1$. And so $\exists\sigma''\forall\sigma'''\ \sigma'' \leq \sigma''' < \sigma\ [\![\phi]\!]^{\delta',M^{\sigma'''}_{+}} = 1$. It follows that $[\![\phi \rightarrow \neg\phi]\!]^{\delta',M^{\sigma'}_{+}} = 0$, and so $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} \neq 1$. A similar argument will show that on the assumption $[\![\phi]\!]^{\delta',M^{\sigma'}_{+}} = 0$, $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} \neq 1$.

So on the assumption that $|||I\phi|||^{\delta'} = 1$ and that there is some $\xi > \sigma\ [\![\phi]\!]^{\delta',M^{\xi}_{+}} \neq 1/2$, it follows that $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} = 1$ and $[\![I\phi]\!]^{\delta',M^{\sigma'}_{+}} \neq 1$. By reductio it follows, on the assumption that $|||I\phi|||^{\delta'} = 1$, that for all $\xi > \sigma\ [\![\phi]\!]^{\delta',M^{\xi}_{+}} = 1/2$. This gives us (D3).

References

Arntzenius, F. (2008). No regrets, or: Edith Piaf revamps decision theory. *Erkenntnis*, 68:277–297. 18

Burge, T. (1978). Buridan and epistemic paradox. *Philosophical Studies*, 34:21–35. 9

Burge, T. (1984). Epistemic paradox. *Journal of Philosophy*, 81(1):5–29. 9

Conee, E. (1982). Against moral dilemmas. *The Philosophical Review*, 91(1):87–97. 18

Conee, E. (1987). Evident, but rationally unacceptable. *Australasian Journal of Philosophy*, 65:316–326. 9

Davidson, D. (1980a). Actions, reasons and causes. In *Essays on Actions and Events*. Oxford University Press. 25

Davidson, D. (1980b). Mental events. In *Essays on Actions and Events*. Oxford University Press. 25

Feldman, R. (2004). Having evidence. In *Evidentialism*, chapter 9. Oxford University Press. 15

Field, H. (2003a). No fact of the matter. *Australasian Journal of Philosophy*, 81(4):457–480. 6, 35, 41

Field, H. (2003b). The semantic paradoxes and the paradoxes of vagueness. In Beall, J. C., editor, *Liars and Heaps*. Oxford University Press. 24

Field, H. (2007). Solving the paradoxes, escaping revenge. In Beall, J. C., editor, *Revenge of the Liar*. Oxford University Press. 4

Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press. 2, 4, 6, 13, 38, 40, A.1, 47, 51

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. 18

Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19):690–716. 1, 6, 12, A.1, 49

Lemmon, E. (1962). Moral dilemmas. *The Philosophical Review*, 71(2). 18

Lewis, D. (1974). Radical interpretation. *Synthese*, 23:331–44. 25

Lewis, D. (1986). A subjectivist's guide to objective chance. In *Philosophical Papers Vol.II*. Oxford University Press. 41

Lewis, D. (1999). Reduction of mind. In *Papers on Metaphysics and Epistemology*. Cambridge University Press. 25

Marcus, R. B. (1980). Moral dilemmas and consistency. *Journal of Philosophy*, 77(3):121–136. 18

McGee, V. (1991). *Truth, Vagueness and Paradox*. Hackett Publishing Company. 24

Osborne, M. and Rubenstein, A. (1994). *A Course in Game Theory*. MIT Press. 18

Parsons, T. (1984). Assertion, denial and the liar paradox. *Journal of Philosophical Logic*, 13:137–152. 6

Priest, G. (2002). Rational dilemmas. *Analysis*, 62(1):11–16. 18

Richard, M. (2008). *When Truth Gives Out*. Oxford University Press. 2, 6

Soames, S. (1999). *Understanding Truth*. Oxford University Press. 2, 6, 24, 47

Sorensen, R. (1988). *Blindspots*. Oxford University Press. 9

Stalnaker, R. (1984). *Inquiry*. MIT Press. 25

Van Fraassen, B. C. (1973). Values and the heart's command. *Journal of Philosophy*, 70(1). 16

Williamson, T. (2000). *Knowledge and it's Limits*. Oxford University Press. 15

Yablo, S. (2003). New grounds for naive truth. In Beall, J. C., editor, *Liars and Heaps*. Oxford University Press. 2