Indeterminacy: an investigation into the Soritical and semantical paradoxes

Andrew Bacon

October 11, 2012

Abstract

According to orthodoxy the study of the Soritical and semantical paradoxes belongs to the domain of the philosophy of language. To solve these paradoxes we need to investigate the nature of words like 'heap' and 'true.'

In this thesis I criticise linguistic explanations of the state of ignorance we find ourselves in when confronted with indeterminate cases and develop a classical non-linguistic theory of indeterminacy in its stead. The view places the study of vagueness and indeterminacy squarely in epistemological terms, situating it within a theory of rational propositional attitudes. The resulting view is applied to a number of problems in the philosophy of vagueness and the semantic paradoxes.

Contents

1 Linguistic and Non-linguistic Theories of Vagueness **1**
- 1.1 Linguistic theories . . . 4
  - 1.1.1 Determinacy operators . . . 5
  - 1.1.2 The paraphrase strategy . . . 6
  - 1.1.3 The revisionary strategy . . . 8
  - 1.1.4 Problems with intensionalism . . . 11
  - 1.1.5 Problems with hyperintensionalism . . . 14
- 1.2 Outline of a non-linguistic theory . . . 15
  - 1.2.1 The theory of propositions . . . 18
  - 1.2.2 Precision operators and determinacy operators . . . 20
  - 1.2.3 Logical issues . . . 22

2 Vagueness and Evidence **27**
- 2.1 Inexact evidence . . . 28
- 2.2 Updating on your evidence . . . 34
- 2.3 The supervenience of vague beliefs on precise beliefs . . . 38
  - 2.3.1 The supervenience thesis . . . 40
- 2.4 A principle of plenitude for vague propositions . . . 41
- 2.5 Further issues and clarifications . . . 43
- 2.6 Appendix . . . 48

3 Vagueness, Uncertainty and Desire **50**
- 3.1 Factualism and non-factualism . . . 51
- 3.2 Vagueness and uncertainty . . . 54
  - 3.2.1 Ersatz uncertainty . . . 55
  - 3.2.2 Rejectionism . . . 57
  - 3.2.3 No uncertainty . . . 60
- 3.3 Vagueness and desire . . . 62
  - 3.3.1 Preferences . . . 62
  - 3.3.2 Reasoning about preferences . . . 63
  - 3.3.3 Representing uncertainty . . . 65
  - 3.3.4 Decision theory . . . 67
- 3.4 May we care intrinsically about the vague? . . . 72
  - 3.4.1 The indifference principle . . . 75
  - 3.4.2 Objections to the indifference principle . . . 76

4 Vagueness At Every Order **78**
- 4.1 The problem of higher order vagueness . . . 80
  - 4.1.1 The problem of higher order vagueness . . . 81
  - 4.1.2 Is a conjunction of determinate truths determinate? . . . 83
  - 4.1.3 Sharp boundaries from B . . . 83
  - 4.1.4 Mahtani on failures of B . . . 84

4.1.5 Sharp boundaries from other principles . . . 84
4.2 A B-free solution . . . 85
4.2.1 Realistic frames . . . 86
4.2.2 Precise ranges . . . 88
4.2.3 The Forced March Sorites and Assertion . . . 89
4.2.4 Other paradoxes . . . 92
4.2.5 Nihilism* . . . 93
4.3 Appendix . . . 94
4.3.1 Further restrictions . . . 96
4.3.2 B entails that a conjunction of determinate truths is determinate . . . 97

5 A Theory Of Indeterminacy 99
5.1 Two examples of indeterminacy . . . 99
5.2 A non-linguistic account . . . 100
5.2.1 Indiscernibility . . . 101
5.2.2 Logical properties of indiscernibility . . . 103
5.2.3 Admissible automorphisms . . . 105
5.2.4 Explanatory priority . . . 109
5.3 The framework . . . 110
5.4 Decision theory . . . 114
5.4.1 The Galois theory of Boolean algebras . . . 116

6 Propositions and Paradoxes 117
6.1 Prior's paradox . . . 118
6.1.1 The paradox . . . 120
6.1.2 The significance of Prior's paradox . . . 122
6.1.3 Variations and refinements . . . 123
6.2 Operators, predicates, sentences and propositions . . . 126
6.2.1 Are sentences and propositions isomorphic? . . . 126
6.2.2 Non-wellfounded propositions . . . 127
6.2.3 Are predicates and operators isomorphic? . . . 129
6.3 Does the liar express a proposition? . . . 130
6.3.1 Liars express propositions . . . 131
6.3.2 Which proposition does the liar express? . . . 133
6.4 A theory of truth and semantic expression . . . 136
6.4.1 A non-disquotational picture of meaning . . . 137
6.4.2 Can a language contain its own truth predicate? . . . 140
6.5 Appendix: A theory of truth and expression . . . 143
6.5.1 Axioms . . . 143
6.5.2 Consistency proof . . . 145

7 Revenge 146
7.1 Linguistic theories and revenge . . . 147
7.1.1 Revenge paradoxes . . . 148
7.1.2 The impossibility of a revenge free linguistic diagnosis of the liar . . . 151
7.2 Non-linguistic theories . . . 155
7.2.1 Unclear truth conditions . . . 156
7.2.2 Higher order unclarity . . . 159
7.2.3 Assertion . . . 161
7.2.4 Can we recover sentential clarity? . . . 162
7.3 A theory of clarity . . . 165
7.3.1 The theory DFS . . . 167
7.3.2 Higher order indeterminacy and revenge . . . 169
7.3.3 Hierarchies of determinacy operators . . . 170

7.3.4 Assertion 172
7.3.5 Assertive uttering 173
7.4 A semantically self-contained language 175
7.4.1 The consistency of DFS 175
7.5 Semantic self-sufficiency 176

List of Figures

2.1 The proposition that Harry is bald neither entails nor is entailed by the proposition that Harry has N hairs 32
2.2 A smooth curve and a sharp curve of n against credence that Harry has n hairs. 33
2.3 A graph of n against the proportion of epistemic states where the tree's height is $n$cm where it's also about 200cm. 37

4.1 Mahtani's counterexample to B 84
4.2 The validity of B^*: x sees a far out point on the edge, yet there is a finite path back to x. 87

5.1 Some common principles and species conditions which suffice for their validity 112

List of Tables

1.1 Vagueness and modality 25

Acknowledgements

Many people, both in print and in conversation, have shaped my thinking on the paradoxes of vagueness and the liar. I would like in particular to thank the graduate students and faculty at Oxford during my time there and audiences at Barcelona, Bristol, Brown, Dubrovnik, Leeds, Munich, Oxford, USC and Yale at which I presented various versions of the ideas developed in this thesis. I received lots of helpful feedback from these people.

Special thanks are due to John Hawthorne and Tim Williamson. Tim has provided me with lots of insightful feedback and comments on my work during the last five years. I thank John for his encouragement and useful discussions on these topics and for taking an interest in my career. I cannot thank him enough for his professional advice and support over the years.

My deepest debt of gratitude is to my supervisor, Cian Dorr, without whose comments and suggestions this thesis would have been something else entirely and much the worse for it. Most of what I know about philosophy I have learned from him.

Thanks of another kind are owed to my fiancée, Helen Davies, without whom I would have probably never finished this thing.

Preface

This thesis is an extended investigation into two classical topics: the paradox of the heap and the paradox of the liar.

We shall be mostly preoccupied with the logical and epistemological puzzles these two topics generate. In both cases the logical aspect of these puzzles takes the form of paradox. Suppose I am adding grains of sand to a pile one by one. At what point does the pile become a heap? The natural thought is that it is not possible for a single grain of sand to make the difference between a heap and a non-heap. The Sorites paradox arises from the observation that, on the assumption that no grain of sand makes the difference, one can prove that either every pile, no matter how small, is a heap, or that no pile, no matter how big, is a heap.

The liar paradox, on the other hand, arises when one considers certain sentences involving self-reference, such as the sentence 'this sentence is not true.' Is it true or isn't it? If it is true, it isn't, since that is what it purportedly says, and similarly one can reason that if it isn't true then it is.

Associated with both these two paradoxes are two epistemological puzzles. Suppose I am pointing at a pile which is neither clearly a heap nor clearly not a heap. Now consider the question

Q1. Is this pile a heap?

Assuming classical logic the pile either is a heap or it isn't. Yet it seems that we don't and couldn't possibly know which, no matter how much evidence we gained about the number grains in the pile, their arrangement, and so on. A somewhat similar phenomenon arises when we are considering self-referential sentences. Instead of considering the sentence 'this sentence is not true' consider the sentence 'this sentence *is* true.' Call this sentence T and ask yourself

Q2. Is T true?

Once again it seems that there is simply no way of determining whether T is true or not and it has nothing to do, it seems, with my impoverished epistemic situation.

According to philosophical orthodoxy both of these problems belong to the philosophy of language. To solve these paradoxes we need to investigate the nature of predicates, like 'heap', and 'true', and it is in revealing the nature of these words that we shall be able to say something illuminating about the logical and the epistemological paradoxes.

Furthermore, both of these antinomies give rise to questions, such as Q1 and Q2, where it is indeterminate, i.e. where there is no fact of the matter, what the correct answer is. In this thesis I criticise the orthodox view and develop a non-linguistic theory of indeterminacy in its stead. The view endorses classical logic and situates the study of vagueness and indeterminacy squarely in epistemological terms, focussing on the role that they play in a theory of rational propositional attitudes.

In chapter 1 I criticise linguistic explanations of why we are unable to know the answer to questions like Q1 and Q2 and lay the basis for what I think a non-linguistic theory should look like. In chapters 2 and 3 I focus on the way in which vagueness interacts with certain

rational propositional attitudes. This raises a number of questions which occupies my discussion: under what circumstances do we acquire vague evidence? Should our degrees of belief be probabilistically coherent? May one care intrinsically about the vague? Do our vague beliefs supervene on our beliefs about the precise? Chapter 2 consists mostly of an extended discussion of the important role that vague evidence plays in epistemology. In chapter 3 we consider the problem of distinguishing the claim that there is no fact of the matter whether p from a purely epistemic claim about our inability to know whether p for certain reasons. To answer this we investigate the attitudes of belief and desire and the roles that vague beliefs and desires play in a theory of decisions. Chapter 4 addresses the paradoxes of higher order vagueness; paradoxes that purport to show that sharp, knowable, boundaries re-emerge in vague predicates at higher orders. In chapter 5 I focus on cases of propositions which are intersubstitutable for one another within a certain class of rational propositional attitudes and use this phenomenon to develop a non-linguistic theory of indeterminacy. Chapter 6 is concerned with a number of paradoxes concerning propositional attitudes due to Prior; I show that these paradoxes are not dealt with by the orthodox linguistic approaches to the semantic paradoxes. Finally in chapter 7 I consider the 'revenge paradoxes', further paradoxes which arise when one tries to solve the liar paradox, and argue that the only way to avoid revenge involves rejecting the widely held assumption that there must be a linguistic distinction between the 'healthy' and 'diseased' sentences or utterances which give rise to the semantic paradoxes.

Chapter 1

Linguistic and Non-linguistic Theories of Vagueness

The view I defend in this thesis departs from much of the literature on vagueness in its rejection of a fairly widespread assumption.

In order to assess this assumption I shall start by developing an analogy with a historically similar debate. Consider the sentence

It's contingent whether Harry is tall. (1.1)

Claims involving modal words, such as (1.1), seem to pose a puzzle; how do facts like this fit into a world built out of protons, neutrons and so on? In order to make sense of this many early analytic philosophers took statements such as (1.1) to be making claims about the way we use language (see for example Ayer [1], Carnap [18], Malcom [88].) Despite its surface form, these philosophers claimed, (1.1) says something about the English sentence 'Harry is tall', perhaps that both its truth and falsity are consistent with current linguistic conventions.

The fact remains, however, that sentences such as (1.1) do not mention sentences. The expression 'it's contingent whether' is grammatically an operator, as are many modal expressions in philosophical usage. It therefore seems that anyone subscribing to a linguistic account of necessity should view these locutions as *dispensible*: either we should be able to paraphrase them away, or we should say that the things we say using them are strictly false and that there are better things to be said in their place using a more perspicuous formulation explicitly mentioning sentences.

Why is the linguistic account of necessity committed to this dispensibility thesis? If, after introducing our notion of a sentence being necessary or contingent by convention, we still have truths such as (1.1) floating about which are not somehow derivative on the conventional use of sentences, it seems we are in no better position than we were originally. We would still need some account of the non-linguistic fact expressed by (1.1), and the puzzles we alluded to concerning their nature are no less worrisome.

Most contemporary philosophers take the view that necessity is not a linguistic phenomenon. There are numerous objections to the linguistic view which I shall not go into (see Sider [121] for a recent discussion) – an unsophisticated yet powerful objection to the view is simply that, unlike its putative paraphrase, the claim expressed by (1.1) seems to be about Harry and his height, and not about any particular sentence of the English language. On the non-linguistic account truths like (1.1) do not need paraphrases. To be sure, there is still a distinction between sentences to be made that is analogous to the one between contingency and necessity; the sentence 'Harry is tall' clearly falls on one side of this distinction while the sentence 'Harry is Harry' falls on the other. For the non-linguistic theorist

this is simply a matter of a sentence expressing a proposition that is contingent or necessary respectively. The operator way of speaking is prior to the linguistic way and there is no further puzzle, for the non-linguistic account, of explaining this difference between sentences.

A debate that is formally exactly analogous to this one arises within the philosophy of vagueness.

The analogy begins with a, if not *the*, concept central to the study of vagueness: the concept of a borderline case. The notion is best introduced by examples. In order to explain what it means to be borderline tall we might begin by pointing out people who are clearly tall, and people who clearly not tall, and then we might assert, once our audience has a feel for distinction, that the remainder, those who are neither clearly tall nor clearly not tall, are the *borderline cases*. There is of course much more to be said: that it's impossible to know whether a borderline case of tallness is tall or not, that properties that could potentially have borderline cases are usually susceptible to Sorites paradoxes, and so on.

Suppose that Harry is borderline tall. We can express this by saying:

> It's borderline whether Harry is tall (1.2)

The property of being borderline tall has in common with the property of being contingently tall the fact that they are usually ascribed using operator like expressions, such as 'it's borderline whether', 'it's vague whether', 'it's indeterminate whether', 'there's no fact of the matter whether', and so on, which do not mention language. When we say that it's vague whether Harry is tall we are not saying something about words or language, we are saying something about Harry and his height, just as when we say that it's contingent whether Harry is tall we are talking about Harry and his height.

Nonetheless, there is clearly a distinction between bits of language to be made that is analogous to the distinction between propositions: the sentence 'Harry is tall' is obviously different from the sentence 'Harry is 1.73 meters' and the difference consists in something closely related to vagueness. For a non-linguistic theory of what a borderline case is this distinction is simply one holding between sentences which express borderline propositions and those which don't.

Perhaps surprisingly, non-linguistic theories of borderlineness are rarely defended.[1] Most philosophers take instead the above distinction between *sentences* to be the basic phenomenon, and choose accordingly to focus their attention on the following linguistic relation (see Dorr [26].)

> Sentence s is borderline as used in the language L by community C in context c (at the the actual world at time t.)[2]

Just to reiterate: it is hardly deniable that this distinction between sentences exists, the question is whether this distinction is basic or whether a sentence is borderline in L as used by C (in c, w and t) purely in virtue of its expressing (in L as used by C at c, w, t) a proposition that is borderline.

The linguistic approach to vagueness, characterised by its preoccupation with the aforementioned relation between sentences, utterances or more generally linguistic items, is shared by most of the most prominent approaches to vagueness. A representative but incomplete list of such approaches include accounts of vagueness that appeal to

- Semantic indecision (Lewis [82], McGee & McLaughlin [92], Keefe [71], Dorr [24], Rayo [113])

[1] With some notable exceptions. See, for example, Fine [43], Field [37], Barnett [4], Barnes and Williams [3].

[2] Why do we need to relativise to both languages and communities? You could imagine that the same community of speakers could simultaneously speak two different languages both of which containing the same sentence (observe the multitude of actual countries that are multilingual). Conversely, the same language may come in different dialects, so that one and the same sentence in the language L can have two different meanings relative to the dialects of different subcommunities of L-speakers.

- Semantic plasticity (Williamson [138], Hawthorne [56])
- Inconsistent meaning constitutive principles (Eklund [29])
- Contextualism (Graff Fara [50], Raffman [110], Shapiro [120])
- Ambiguity (Sider and Braun [15])
- Use theories of meaning (Horwich [61])
- Nihilism about vague language (Unger [132])

Philosophers subscribing to the linguistic approach appear to be committed to a version of the dispensibility thesis we discussed in relation to modal language. Either claims involving operator like expressions, such as (1.2), are paraphraseable in terms of a sentential vagueness predicate, or these claims are simply false and some revision to our language is required.

In order to make this commitment vivid, it's worth pointing out that two of the most fundamental issues in the philosophy of vagueness do not mention language or a linguistic vagueness predicate: the problem of explaining why we cannot rationally know whether someone is tall (say) if it's borderline whether they're tall, and the Sorites paradox. To be sure, there are linguistic Sorites: for example, just as there's a Sorites for tallness, there's a Sorites for the complex property of being such that the word 'tall' applies to you in English. But this latter Sorites looks like it at best demonstrates the vagueness of the applying relation since the name 'tall' is a precise name for a word. Similarly, due to the truth of its consequent, it's true that (as it happens, due to the way Germans use German) if the sentence 'Harry ist hochgewachsen' is vague in German it's impossible to find out whether Harry is tall, but this fact would be easily explainable given an answer to the original question of why we cannot rationally know whether Harry is tall if it's borderline whether he's tall without appealing to facts about the German language.

So why are these philosophers committed to something like the dispensibility of operator-talk? In the case of contingency the thought was that if 'operatorese', locutions like that in (1.1), are indispensible to our way of talking about modality then anything we've said about linguistic modality has done little to explicate (1.1) or elucidate the puzzles concerning the nature of modality. An analogous point applies here: if operatorese is indispensible to the way we describe at least *some* of the standard puzzles involving borderline cases or Soritese sequences, then the distinction between kinds of sentences is at best an interesting side issue, and the central problems in the philosophy of vagueness remain standing. We must still address the Sorites paradox for tallness (as opposed to the Sorites paradox for being such that 'tall' applies to you in English) and the problem of explaining why we cannot know whether Harry is tall (as opposed to the problem of explaining why we cannot know whether the sentence 'Harry is tall' is true in English.)

A failure to accommodate the Sorites paradox and the puzzle about ignorance, or at least relate them to one's theory of vagueness, would constitute a serious lacuna in any account of vagueness. So whether or not I am right that the majority of philosophers working on vagueness are committed to the dispensibility thesis, I have done enough to motivate the primary goal of this thesis which is to fill this lacuna: if operatorese is not eliminable in favour of a linguistic distinction, then we need to say something about the Soritical and epistemological puzzles stated in operatorese; in other words we need a non-linguistic account of operatorese. This is what I shall do in this thesis. Many of the philosophers listed offer an analysis of what it is for a sentence to be vague in a language, relative to the relevant parameters, which does not appeal to the vagueness of the proposition that sentence expresses. It is therefore compatible with the indispensibility of operatorese that both the distinction between sentences and the distinction between proposition are basic and exist in parallel. But once one has accepted the basicness of operatorese, the converse dispensibility thesis becomes very attractive: that a sentence is vague in L (relative to ...) if it expresses in L (relative to ...) a borderline proposition.

1.1 Linguistic theories

It will be useful, at points, to distinguish two ways one might be a linguistic theorist. I shall call them the 'intensionalists' and the 'hyperintensionalists' accordingly. The difference concerns whether one accepts the existence of vague properties and propositions. However this makes the thesis sound more controversial than it is – by a hyperintensionalist I mean anyone who takes the following argument to be sound:

> For each n, one can (and often must) rationally believe that Harry has less than n hairs without believing that Harry is bald or believe that Harry is bald without believing that Harry has less than n hairs.
>
> Therefore, for each n, the proposition that Harry is bald is distinct from the proposition that Harry has less than n hairs (Leibniz's law, reasoning by cases.)

This reasoning extends to any precise proposition about Harry's head. In order to keep the discussion clean, however, I shall work under the simplifying assumption that whether someone is bald supervenes on the precise number of hairs they have.[3]

An alternative view does not single out vagueness as a special source of hyperintensionality. The simplest form of this kind of view identifies propositions with sets of possible worlds, although similar points apply to less radical versions of this view.[4] I shall call this the intensionalist account of propositions.

The intensionalist view does not postulate a further proposition – the proposition that Harry is bald – distinct from the propositions of the form: that Harry has less than n hairs.[5] Under the supervenience assumption the intensionalist, in contrast to the hyperintensionalist, would identify the proposition that Harry is bald with the proposition that Harry has less than n hairs, for some suitable choice of n. Thus the intensionalist needs to say something about the above argument which appears to show that these propositions are distinct. A fairly standard response for intensionalists, often appealed to in the literature on proper names, is to deny that the conclusion is an instance of Leibniz's law despite its apparent surface form. Although the response can be manifested in various different ways, the kernel of the idea is that sentences like 'A believes that p' do not express binary relations between people and propositions, but ternary relations between people propositions and something else: $Bel(A, p, x)$. The most we can infer from $Bel(A, p, x)$ and $\neg Bel(A, q, y)$ by Leibniz's law is that if $p = q$ then $x \neq y$. Since the third argument is not part of the surface form it is often assumed it is supplied contextually.[6] It will be useful to have an easy to pronounce word for the relation $Bel(A, p, x)$ – I shall say 'A believes* p relative to x'; I shall follow this convention of starring with other propositional attitudes as well.

What kind of thing goes into the third argument place and what does it mean for a person to be related to a proposition and one of these things? Here is a very schematic picture of what is going on: in order to have a belief that p, one must be in the right kind of belief state. However, for an intensionalist about propositions there are typically a large variety of fundamentally different belief states can result in a belief with the same content. One can believe that p in many different ways by having mental tokens of different representations. I can, for example, have the belief that Hesperus is Phosphorus very easily by producing a token mental representation of something fairly trivial – the mental

[3] I am implicitly assuming that the vague facts supervene on the precise facts.

[4] One might think this identification is an overkill; however it makes the idea that precise sentences and vague sentences express the same kinds of proposition particularly vivid.

[5] This fact follows from classical logic, the simplifying assumptions we made about the supervenience base for baldness facts, and the principle that necessarily equivalent propositions are identical: by classical logic and the supervenience assumption there's a last number of hairs such that having that many hairs necessary and sufficient for being bald. So the proposition that Harry is bald is necessarily equivalent, and thus identical, to the proposition that Harry has less than that number of hairs.

[6] Views which build the extra argument into the propositions – for example, by identifying propositions with pairs of sets of worlds and guises – count as hyperintensional on this classification.

equivalent of the sentence 'Hesperus is Hesperus' for example. And I can have a belief with the same content by tokening a less trivial mental representation, the mental equivalent of the sentence 'Hesperus is Phosphorus' for example. The third argument is therefore supposed to be filled by a representation of some kind, and the relation expressed by a belief attribution holds between a person, representation and proposition, p, only if the person is having a mental token of that representation and is thereby having a belief that p. Let us introduce the name 'guise' for whatever kind of representation plays the role just described. In short, the view just described is one in which one and the same proposition may be known and unknown relative to different guises.

1.1.1 Determinacy operators

Intuitively we can't know whether Harry is bald (unless he had a different number of hairs), we shouldn't assert that he's bald or question whether he is, because there simply is no fact to be known, or fact to be enquired after. A natural and also common way to model this way of talking is to introduce an operator, Δp, pronounced 'it's determinate that p', and to introduce or define an operator ∇p pronounced 'it's indeterminate whether p', or 'there's no fact of the matter whether p' to distinguish between scenarios in which the subject matter is factual and those in which it is not. If we are interested in vagueness, a species of indeterminacy, we could introduce analogous operators. For most applications we don't need to know too much about what these ascriptions of indeterminacy or vagueness amount to, only what their logical properties are, how they constrain knowledge, the role they play in the pragmatics of assertion and questions, and so on. I shall call principles like these aspects of the 'determinacy role'. If I am right, the determinacy role is independent of one's particular analysis of indeterminacy (or vagueness), and encodes principles that govern the basic use of the notion and which allow it to do explanatory work.

The non-linguistic theory I defend in this thesis employs an operator which I show will obey the determinacy role described below. An operator satisfying the determinacy role is well placed to explain a number of the paradoxes and issues associated with borderline cases.

Logical The operator 'it's determinate that p' is governed by the modal logic KT, or some normal extension of it:

Closure If it's determinate that if p then q, and it's determinate that p, then it's determinate that q.

Factivity If it's determinate that p, then p

Necessitation If one can prove p from classical logic and these three principles, one can prove that it's determinate that p.

Epistemic If it's indeterminate whether p, then it's not rationally known whether p.

Doxastic No rational person who is certain that it's indeterminate whether p can be rationally certain that p or rationally certain that $\neg p$.

Pragmatic

Assertion It is not permissible to assert that p (that $\neg p$) if one believes that it's indeterminate whether p.

Questions It is not permissible to ask whether p if you have been told that it's indeterminate whether p.

Bouletic It is not rational to care intrinsically about the indeterminate.

Attitudes It is not rational to wonder whether, hope that, fear that, wish that, [...] p, if you believe that it is indeterminate that p.

This represents an important bundle of principles that my own theory will provide. It will be useful to bear these in mind for comparison in what follows. However, it is clear that many of these principles are controversial. That said, many philosophers have assumed or defended theories that support **Logical**, **Epistemic**, **Doxastic** and **Pragmatic**, and the latter three are often taken to be data which a theory of vagueness is supposed to explain.

Note that the determinacy role is closely connected to the operator approach; when we say we cannot know (rationally believe, fear, hope, etc) that p because it is indeterminate whether p, we can do so because the objects of ignorance, uncertainty and other propositional attitudes are the same as the objects of indeterminacy. This approach to the logic and epistemology of vagueness, call it 'the operator approach', is widely adopted in the literature on indeterminacy, and for good reason since operators like these are very well understood (see Blackburn et al. [10].)

However it is unclear whether a linguistic theory of vagueness can play the role we've outlined above. According to linguistic accounts of vagueness sentences or utterances are usually the bearers of vagueness. Perhaps we can accommodate the thought encoded by **Pragmatic** by replacing the operator 'assert that' with the sentential predicate 'assertively utter'. However the other cases are not so easy; in order for indeterminacy to play **Epistemic**, **Doxastic**, **Bouletic**, **Attitudes** we must posit contentious principles connecting the possibility and rationality of knowledge and other propositional attitudes to the status of particular sentences in particular languages. The case of **Logical** is also difficult as a straightforward translation of KT to a sentential predicate is inconsistent with a theory rich enough to describe its own syntax [95].

It might seem hard to imagine that matters concerning the *role* of vagueness could turn on whether vagueness is treated as a linguistic phenomenon or not. Surely, one might think, it is possible to paraphrase anything you can do in the operator way using only vocabulary a linguistic theorist accepts. Call this the 'paraphrase strategy'.

I'll begin by showing that this thought is mistaken; it is not possible to simply paraphrase statements of the standard epistemological and logical role using a linguistic predicate. Then, in §1.1.3, I'll show that nothing even resembling the determinacy role can be recovered in a linguistic framework. There are simply no plausible epistemological principles governed by vagueness conceived as a linguistic property.

1.1.2 The paraphrase strategy

Let us begin by considering the paraphrase strategy. Consider a typical example of something we would state using an operator:

> It's vague whether Harry is bald. (1.3)

To paraphrase (1.3), we must pick a language, L, and sentence, s, such that the claim that s is vague in L (at context c and other parameters) is a reasonable paraphrase of the claim expressed by (1.3); one of the things this paraphrase must do is play the same role that (1.3) does within the determinacy role. An example of the paraphrase strategy, therefore, would be to choose English as our language, and the sentence "Harry is bald". Our paraphrase of (1.3) is thus that the sentence "Harry is bald" is used by English speakers in whatever way is required to make it a vague sentence of English. This is essentially the strategy adopted in Dorr [26]; there the strategy seems to be to simply compose the quotation subnective (of type te) with borderlineness predicate (of type et) to produce an operator (of type tt.)

However there are a number of issues with this paraphrase. An obvious objection is that it is not equivalent to (1.3): English speakers could have used "Harry is bald" in a precise way, to mean that 1+1=2 say, yet it would still have been vague whether Harry is bald provided he has the same number of hairs he in fact has. The equally obvious reply is that we should instead paraphrase (1.3) with the claim that "Harry is bald" is a vague sentence of English *as it is actually used.* But even this seems inadequate: a non-English

speaker may not know anything about how English is actually used, but still know that the number of hairs Harry has falls within the borderline region for baldness. Thus one can know whether Harry is borderline bald without knowing whether the sentence "Harry is bald" is vague as it is actually used in English. Conversely, someone who doesn't speak English may have it on good authority that the sentence "Harry is bald" is vague in English as it is actually used, but have no idea whether Harry has 0 hairs, 1,000,000 or a borderline number, since she might not know what "Harry is bald" actually means in English.

Another issue is that it's totally unclear why the claim that "Harry is bald" is vague in English should play the epistemic, doxastic, bouletic or attitudal roles described – especially for non-English speakers (we shall come back to this later.) If it were a genuine paraphrase of (1.3), however, it would play these roles.

A more pedestrian worry is the dependence of the paraphrase on the choice of sentence and language: the claim that "Harry is bald" is vague in English, and the claim that "Harry ist kahl" is vague in German are two completely different paraphrases of (1.3). One can be true while the other is false – they state completely different facts about different sentences in different languages. But as far as I can see, neither paraphrase is superior to the other – if two incompatible paraphrases are equally good, neither are adequate. On the other hand, what is asserted by (1.3) can be said in a number of languages without mentioning English. The operator formulation does not seems to be about sentences, the English language or the German language, it seems to be about Harry and the status of his head.

Perhaps we could do better with language independent paraphrases:

E. There is some sentence, s, in some language, L, whose linguistic community actually uses s (i) in such a way that it says that Harry is bald and (ii) in whatever way it takes to make s vague in L .

U. Every sentence, s, in any language, L, whose linguistic community actually (i) uses s in such a way that it says that Harry is bald (ii) uses s in whatever way it takes to make s vague in L .

These paraphrases still mention languages and sentences, unlike (1.3), but perhaps they do better with regard to our first problem of language dependence.

Unfortunately, according to some linguistic theories (i) and (ii) are incompatible: the vagueness of a sentence relative to its use in a linguistic community precludes that use determining any proposition as being uniquely expressed by s in L.

However, I think the most problematic feature of this strategy is that it's either inadequate or it assumes the distinction between propositions we are trying to dispense with. Consider the following two possibilities.

> The proposition that electrons are positively charged is expressed in some language by a vague sentence.
>
> The proposition that Harry is bald is expressed in some language by a precise sentence.

If the first claim is possible then E. would be inadequate and if the second statement was possible U. would be inadequate. (It should be noted straight off the bat that intensionalists are already committed to both these possibilities: for example, for some numeral 'N', the precise sentence 'Harry has less than N hairs' expresses the proposition that Harry is bald. The paraphrases E. and U. are simply not available to the intensionalist in the first place; what follows is therefore aimed at hyperintensionalists.)

So it follows that if the paraphrase is adequate neither statement is possible. But facts like this call out for explanation! What is so special about the proposition that Harry is bald that means it can't even *possibly* be expressed by a precise sentence? Yet it seems hard to see how the difference between propositions which could be expressed by a precise sentence and those which couldn't could be explained without appealing to the propositional notion of vagueness that we are attempting to eliminate.

To put it another way, while we know that the proposition that Harry is bald is expressed in English by the vague sentence "Harry is bald" and in German by the vague sentence "Harry ist kahl", what is it about *the proposition* that Harry is bald, that prevents it being expressed by a precise sentence? Couldn't there be a language with a completely precise predicate, such that there is no unclarity about when it applies in that language and when it doesn't apply, which is such that it applies in L only when the object is bald, and doesn't apply otherwise. How are we supposed to distinguish the proposition that Harry is bald from a proposition that can be expressed by a precise sentence, like the proposition that Harry has less than 2,000 hairs, if not by employing the distinction between vague and precise propositions we are trying to eliminate? It seems like we need something similar to the operator way of talking to distinguish vague from precise propositions. The upshot of the possibility of the proposition that Harry is bald being expressed by a precise sentence is that (1.3) may come apart from its paraphrase.

1.1.3 The revisionary strategy

The prospects of providing paraphrases of claims like (1.3) look dim. I think the best strategy for a linguistic theorist to adopt is to stop talking in the operator way altogether, and to only ascribe vagueness using a predicate applying to linguistic items. This line of response is genuinely revisionary: we must simply reject the determinacy role, as stated, and offer alternative explanations of the various facts that we explained using the determinacy role. For example, the fact that we cannot know whether Harry is bald still needs explaining, we need to be able to explain this in a way that does not appeal to the principle **Epistemic**. We must revise the logical role indeterminacy is usually taken to assume, and we must revise the epistemological, doxastic, bouletic and attitudinal role so that it is governed by a relation between sentences and languages and not an operator or predicate applying to propositions.

However there are immense technical obstacles in the way of carrying out such a project. To give the flavour of the kinds of problems one might face, note that one cannot have the analogue of the logical role for a determinacy predicate, $Det(x)$, since the principles $\vdash Det(\ulcorner\phi\urcorner) \rightarrow \phi$ and rule of proof that if $\vdash \phi$ then $\vdash Det(\ulcorner\phi\urcorner)$ are inconsistent in any system with a suitably rich theory of syntax (see Montague [95].)[7] It is thus unclear whether one can revise the logical role of indeterminacy in a way consonant with the way it is used in the rest of philosophy.

Whether or not these technical difficulties can be overcome in a way that does not prevent one from saying everything one wants is a matter for debate. But there are other obstacles for the project of revising the determinacy role: our beliefs about indeterminacy play an important role in regulating our other propositional attitudes. But, I shall argue, any revision of **Bouletic**, **Epistemic**, **Doxistic** and **Attitudes** that are instead regulated by beliefs about the properties of a sentence in a particular language will be false.

Explaining particular unknowable propositions

Suppose that Harry is borderline bald. Many philosophers have taken it for granted that we do not and cannot know whether Harry is bald, and that explaining this fact is one of the most important tasks in the philosophy of vagueness. In this particular situation the relevant fact that needs explaining is

> **Ignorance**: it is not possible to rationally know whether Harry is bald.

It seems that we simply couldn't find out whether Harry is bald – at least, not unless Harry had a different number of hairs. No evidence could be obtained, whether that be counting

[7] This problem is particularly pertinent when we appeal to indeterminacy to explain the fact that we cannot know whether the truth-teller is true, which are closely related to the semantic paradoxes.

all the hairs on his head, or reflecting on the nature of baldness, that would put us in a better epistemic situation with regard to Harry's baldness.

Ignorance is an instance of one of a tightly knit cluster of phenomena that we associate closely with borderlineness. There are systematic connections between facts like **Ignorance** and borderline cases – facts like this just call out for some kind of general explanation.

According to the determinacy role, that explanation takes the form of a general principle, which in turn falls out of a general theory of indeterminate propositions:[8]

> **Epistemic** Necessarily, if it's borderline whether p, then it's not rationally known whether p.

A linguistic theorist who has no paraphrase for the operator locution employed in **Epistemic** cannot accept it as an explanation. If she is to offer a general explanation of **Ignorance** which still has something to do with vagueness she must instead seek to explain **Ignorance** in terms of her favoured vocabulary: s is borderline in L relative to context c and other parameters. But this raises a number of questions

> The fact that some sentence, s, is borderline as used in language L, by community C, at context c, at time t and world w purportedly explains **Ignorance**. We must ask
>
> (a) Which language L and linguistic community?
>
> (b) Which sentence of this language is such that its borderlineness explains **Ignorance**?
>
> (c) At what time must the sentence be borderline?
>
> (d) What context must the sentence be borderline in?
>
> (e) At what world must the sentence be borderline?

Which language should we choose? Could the linguistic practices of people in China, for example, prevent me from knowing whether Harry is bald? It seems to me to be obvious that they could not. Perhaps we can only explain *my* ignorance by appealing to the borderliness of an English sentence, a native Spanish speakers ignorance by appealing to the borderliness of a particular Spanish sentence, and so on. But this is not a particularly good explanation after all, due to its lack of generality. Not to mention its lack of plausibility! What if I don't speak a language, for example? Would that make it easier for me to find out whether Harry is bald? Even if I do speak a language, why should my co-speakers linguistic habits bear on whether I can find out whether Harry is bald?

Which sentence ought to be used to explain **Ignorance**? It can't be, for example, the borderliness of the sentence 'John in tall' – presumably it must be a sentence which expresses the proposition that Harry is bald in L. What if there is no sentence in my native language expressing the proposition that Harry is bald? This would not make it easier for me to find out whether Harry is bald. This would be particularly troubling for a hyperintensionalist. Furthermore, what if, as some theories claim, vagueness *prevents* a sentences expressing a unique proposition? Must the proposition that Harry is bald be merely *among* the propositions s expresses, or among the propositions s doesn't determinately fail to express?

Finally, it seems we must specify a time, world and context at which the sentence is borderline. Words often become more precise as language evolves. If the selected sentence s ('Harry is bald' say) had once been precise it would presumably still be impossible to know whether Harry is bald. Maybe s has to be vague at the time I'm trying to find out whether Harry is bald. If we have to pick a different sentence each time we want to explain why someone can't figure out whether Harry is bald then the explanation loses its generality.

Furthermore, had we used s in a precise way we wouldn't be in a better epistemic situation regarding Harry's head. We may therefore be able to explain why no-one *in fact*

[8] Note that not every theorist has agreed with this intuition. For example in recent years Dorr [24] and Barnett [6] have argued that we can know in indeterminate cases.

knows whether Harry is bald by appealing to a selected sentences borderliness, but in order to explain why we *couldn't* have known whether Harry is bald, even if s had been used in a precise way, we need also to relativise to worlds. This opens up the possibility that we can't know that Harry is bald because the sentence 'John is tall' is used in a vague way at another world to mean that Harry is bald.

Let me now turn to two specific attempts to explain **Ignorance** within a linguistic theory.

The first arises from the epistemicist account of vagueness defended by Williamson [138]. Williamson's view as I am characterising it is linguistic: a sentence is vague relative to a linguistic community C and parameters if, roughly, there are close worlds, with respects to how that language is used, where that sentence says something false and close worlds where it says something true. It earns the name 'epistemicism' as it allegedly *entails* the ignorance thesis. If this is true it is a significant benefit of the view over other linguistic theories. However whether it entails the ignorance thesis is disputable, for it relies on a controversial 'metalinguistic safety' principle.[9] It is unclear how the use of a certain sentence by English speakers at closeby worlds could have a bearing on whether I know that Harry is bald, a claim ostensibly about Harry and the status of his hairline, and not at all about English speakers or the English language. Metalinguistic safety principles have come under significant criticism in recent years (see Magidor and Kearns [70], Sennet [119].)

Note that we should distinguish **Ignorance** from the claim that it is impossible to know whether the sentence 'Harry is bald' is true in English in 2012, at parameters (...). The fact that there are close worlds where we use 'Harry is bald' differently may explain, by *ordinary* safety principles, why we cannot know that this sentence is true in English. But whether Harry is bald or not is not unstable in the same way: he has the same number of hairs (or at least a number within the borderline interval) at all the relevant close worlds. So there is also the further fact that, given the status of Harry's head, we cannot know whether he is bald, and there is no obvious way to infer ignorance of the latter fact from ignorance of the former linguistic fact. In general it is quite easy, if you don't know what certain English sentences mean, to know that Harry is bald without knowing that 'Harry is bald' is true in English, and to know that 'Harry is bald' is true in English without knowing that Harry is bald. Why shouldn't the kind of ignorance about what sentences mean, on the Williamsonian picture, be the same as the kind of ignorance about meaning non-English speakers have?

The second putative explanation appeals to the fact that when there is widespread disagreement about whether a term applies in given case one ought to withhold judgement about whether it applies. One might attribute uncertainty about whether a given vague predicate applies in a particular case to uncertainty about how other people in your community are using that term in the language. While these explanations certainly may show why we can't know whether the predicate 'bald' applies to Harry in English, once again there is the further fact that we cannot know whether Harry is bald and this is not explained by the former facts.

Let me end by mentioning one more point which, although not a fully fleshed out objection, is something I find worrisome. In many ways ways explaining **Ignorance** is one of the easier jobs for a linguistic theorist. Other facts are harder to explain. Suppose that Harry is as before a borderline case of baldness, and that we rationally believe this. Then it seems that

> **Agnosticism**: It would be irrational to believe that Harry is bald, and it would be irrational to believe that Harry is not bald.

This fact is *prima facie* quite puzzling. After all you know that either Harry is bald or he isn't, so at least one of the two beliefs above is true. Furthermore you have all the evidence

[9] For further discussion of metalinguistic safety principles see Magidor and Kearnes [70] .

you could possibly have available. It feels like there should be some general fact about vagueness related phonemona that explains this.

Since there appear to be plenty of things we do not *know* which are not irrational to believe, it seems, therefore, that there is a separate and harder problem of explaining what feature of linguistically vague sentences prevents us from rationally believing propositions. This seems a lot harder to do. For example, in a discussion of Williamson's epistemicism, Horwich writes: "the ignorance due to vagueness is attributed to a special form of unreliability – an external failure – and not, as it should be, to the internal difficulty in making a judgement." [61]. But surely any explanation of our ignorance that appeals to the use of a term in a public language is going to be an external one.[10] An account of non-linguistic indeterminacy that satisfied the determinacy role could explain, in a general way, **Agnosticism** from **Doxastic**, but I am much less confident that an explanation of **Agnosticism** in terms of language use could be given.

1.1.4 Problems with intensionalism

An intensionalist appears to be in a slightly better situation than a hyperintensionalist when it comes to providing a general explanation of **Ignorance**. For example, they would be in a better position under the following assumption:

M Every public language sentence s of $\mathcal{L}$ is intertranslateable with a mental representation $m(s)$

the intertranslateability assumption is left unspecific for the time being.[11] Assume also that for every sentence s of $\mathcal{L}$ there is a sentence expressing the negation of what s expresses in $\mathcal{L}$, $neg(s)$, and that s is vague iff $neg(s)$ is.[12] We might then be able to explain **Ignorance** by appealing to and defending a general principle of the form:

> **The Linguistic Ignorance Thesis** Necessarily, if s is indeterminate in $\mathcal{L}$, then for no proposition p does any rational person know* P relative to $m(s)$ or know* p relative to $m(neg(s))$.

The upshot of this, speaking loosely, is that it may indeed be possible to know, under certain guises, that Harry is bald even when 'Harry is bald' is indeterminate. However, it is not possible to know that Harry is bald relative to any guise corresponding to a sentence which is indeterminate in $\mathcal{L}$.

There is still the task of explaining why the ignorance thesis holds: i.e. the task of explaining why, according to one's preferred account of linguistic indeterminacy, a sentences indeterminacy in a language implies the unknowability of any proposition relative to that sentences corresponding guise. It is unclear to me how this might go on particular proposals, such as semantic indecision, contextualist or inconsistency views, however the prospects for the intensionalist seems, *prima facie*, slightly more promising than in the case where there is an independent non-linguistic truth, Harry's baldness, ignorance of which needs explaining in purely linguistic terms.

It is worth pointing out that this strategy of relativising propositional attitudes will have to extend beyond knowledge and belief if we are to keep other aspects of the determinacy role. For example, the ignorance thesis constrains the assertability of certain propositions, which might be derived from norms of assertion of the following kind:

[10] Horwich's own view may be an exception: Horwich, unlike most linguistic theorists, takes the language in question to be an internal language of thought instead of a public language. This exception aside, the problem of explaining **Agnosticism** seems harder for most linguistic theories.

[11] It is natural to assume things like 'one should not assertively utter s in $\mathcal{L}$ unless one has a token of $m(s)$ in one's belief box' and 'knowing an utterance of s is knowledgeable in $\mathcal{L}$ licenses one to have a belief token of $m(s)$', and so on.

[12] According to the intensionalist there will be both precise and vague sentences which express the negation of what 'Harry is bald' expresses. This assumption might be more accurately expressed by the claim that the language has a negation operator which, when prepended to a vague sentence, produces a vague sentence.

A You may assert that p only if you know that p (justifiably believe that p, et cetera.)

Note that according to the view there are contexts according to which 'I know that Harry is bald' is true, even if Harry has a critical number of hairs. If *A* holds in these contexts it follows that 'you may assert that Harry is bald' is true in these contexts. However in normal contexts 'you may assert that Harry is bald' is false when Harry has a critical number of hairs. In order to account for contexts of both kinds we must postulate context sensitivity in the sentence 'you may assert that Harry is bald'. It is natural to think that the word 'assert' is context sensitive, and that the correct rendering of *A*, stated in non-context sensitive vocabulary, is: *A* may assert* that p relative to s only if *A* knows* that p relative to $m(s)$, where one has asserted* that p relative to s only if one has asserted p in virtue of having uttered s.

Similar reasoning shows that a wide variety of propositional attitudes should be relativised:

S If you have seen that p then you know that p

Suppose that 25m is the critical height it takes for a tree to be tall (ignoring, temporarily, the context sensitivity of 'tall'.) Let's suppose that I have seen that the tree 25m – for example, let's say I can see that it is the same height as a tree whose height I already know, via measurement, to be 25m. Since that the tree is 25m is the same proposition as the proposition that the tree is the critical height for being a tall tree, it follows via Leibniz's law style reasoning that I've seen, and therefore know, that the tree is is the critical height for being a tall tree. However, since *S* seems to express a context insensitive truth, and '*A* knows that the tree is the critical height for being a tall tree' expresses a falsehood in most normal contexts, it seems that '*A* has seen that the critical height for being a tall tree' must also be context sensitive. Again, the most straightforward way of guaranteeing that things work out nicely is to have perceptual reports such as '*A* sees that p' expressing ternary relations between people, propositions and guises.

Problems for intensionalism

While the intensionalist strategy evades some of our concerns it is a fairly radical thesis. For example, let us suppose that I know that Harry's hairline has receded to the point at which it is borderline whether he is bald or not bald. Let us further suppose I know exactly how many hairs he has, namely N, and that N is below the cut-off for being bald. Now, even though Harry is, as it turns out, bald, the following claims seem to be clearly false in any context in which they are uttered:

> I know that Harry is bald.
>
> I may assert that Harry is bald.
>
> I can see that Harry is bald.
>
> I discovered that Harry was bald by counting the hairs of his head.
>
> It is always possible to find out whether someone is bald by counting the hairs on their head.

It is a serious cost that the intensionalist predicts contexts in which utterances of the above sentences are true. Perhaps the intensionalist could argue that, although there are contexts in which the above are true, there are no contexts in which it is appropriate to utter them, and this explains our intuitions about these cases. For example, if you were to utter 'I know that Harry is bald' this would be an inappropriate utterance to make, even if true, because you should in general strive to assert a proposition via a guise that you believe the proposition relative to. Suppose v is a guise such that I know* that Harry is bald relative

to v. You may assert the true proposition that you know* that Harry is bald relative to v via the vague guise corresponding to (this particular utterance of) 'I know that Harry is bald', only if you know, via that same guise, that you know* Harry is bald relative to v. This can all be explained by the fact that the contexts in which 'I know that Harry is bald' are true are also contexts relative to which the sentence is linguistically indeterminate.

Unfortunately this strategy for explaining the bizarreness of the claims above has a limited application. For example it does not extend to simple variants of the above sentence: 'Fred knows whether Harry is bald', 'if Harry is bald Fred may assert that he's bald, and if he isn't Fred may assert that he isn't', and so on. For example, if the context provides a guise, g, such that Fred knows whether Harry is bald relative to it, then 'Fred knows whether Harry is bald' is a determinate sentence in that context, in this context it expresses the proposition that Fred knows* whether Harry is bald relative to g and also a sentence such that I may know* that Fred knows* whether Harry is bald relative to g relative to the guise m('Fred knows whether Harry is bald') (since m('Fred knows whether Harry is bald') corresponds to a determinate sentence in this context.)

Another worry with the approach is that it seems not to account for one of the fundamental problems in the philosophy of vagueness: the Sorites paradoxes. Intuitively there is distinction between two kinds of predicates, vague predicates which are Soritesable – can be inserted into the premises of a Soritese argument all of whose premises sound plausible – such as the predicate 'is bald', and those which are not, such as 'has less than N' hairs. According to the orthodox operator way of speaking, a property, F, is vague just in case it could have had a borderline case, in other words, just in case it's possible that for some x it's vague whether x is F; we have a clear connection between locution 'it's vague whether' and concept of predicate vagueness that arises from the Sorites paradoxes.

It is not clear what it means for a predicate to be vague for a linguistic theorist purely in terms of a sentential vagueness predicate. To say that a predicate, 'F', is vague just in case for some name, 'a', 'Fa' is linguistically vague will not do – 'is 29,000 feet high' is not a vague predicate even though 'Mt. Everest is 29,000 feet high' is linguistically vague. Unlike the hyperintensionalist, the intensionalist may mimic the operator approach by using the guise relative vagueness operator: a property, F, is vague iff for some x it's vague whether x is F relative to a. Yet everything depends on the choice of guise, a – we cannot simply existentially quantify it out for the reasons just mentioned.

A third issue for the intensionalist is the use of the assumption M: that there is some mental guise corresponding to each public language sentence. This encodes a controversial empirical assumption: that there is a language of thought, and that every public language sentence corresponds to a sentence in the language of thought. One might think that an adequate explanation of **Ignorance** should not rest on controversial empirical assumptions like this.

Even more troubling, it seems to make the possibility of having a belief that Harry is bald, or indeed any other vague thought, parasitic on one's bearing the right kind of relation to a vague public language (presumably one containing a sentence synonymous with 'Harry is bald'.[13]) This point can perhaps be seen most vividly with vague beliefs which we seem to be able to acquire in a way that seems to have nothing to do with language, and whose acquisition does not appear to depend on one's ability to speak a public language. For example, if I look at Harry I gain certain information about the number of hairs on his head. In doing so I see a number of things about his head; I've seen that he has more than 0 hairs and less than 10^6. The strongest proposition I have seen, and the strongest belief I have thereby acquired, however does not seem to be a precise belief such as the belief that Harry has between 2,459 and 3,056 hairs. It seems far more likely to me that the strongest belief I have acquired (or at least, have justification for) is a vague belief such as the belief that Harry is bald; and analogously that I've had a vague perceptual experience of seeing that Harry is bald. The intensionalist will cash out this talk of a 'vague belief' and 'vague

[13]'Synonymous with' here cannot simply mean 'expresses the same proposition as.'

perceptual experience' in terms of my mentally tokening some guise which corresponds to a linguistically indeterminate public language sentence. But what public language sentence could possibly correspond to the kind of perceptual experience I have in the case described? Furthermore it does not seem plausible that I need to speak a public language, or be related in any particular way to one, in order to have acquired a vague belief perceptually. This point is not limited to visual capacities; in the next chapter I shall argue that most of our perceptually acquired beliefs, gained through hearing, smelling, proprioception, as well as beliefs gained via offline imprecise calculations about the movement and position of objects, are vague. Only a small proportion of our vague beliefs, those acquired by being told something for example, have any clear cut relation to language.

Let me briefly mention two more worries before we move on. As we mentioned already, there seems to be a separate and harder problem of explaining **Agnosticism**. It is not clear what feature of linguistically vague sentences prevents us from rationally believing propositions, via their corresponding guises, or otherwise, even if we may believe the same propositions via guises corresponding to precise guises.

The second worry is based on the observation that, in some sense or other, what beliefs we have partly explain how we act. On the view we are considering, however, such explanations cannot be carried purely by appealing to the contents of these beliefs – they will have to involve the guises we employ in order to have those beliefs, otherwise we could not explain the difference in behaviour between people with necessarily equivalent beliefs. It becomes harder to see how we might get a systematic connection between beliefs and action of the kind you get from decision theory, for example.

Let us introduce a four place relation, credence*, between people, propositions, numbers and guises: we'll pronounce this 'A's credence* in p relative to guise a is x.' Note that, relative to different guises, you can have numerous different credence*s towards one and the same proposition. For example, my credence* towards the necessary proposition will change depending on whether I am having it in virtue of the guises corresponding to 'Hesperus is Phosphorus', 'every even integer greater than 2 is the sum of two primes' and so on. On the other hand, I can have credence* 1 in p relative to some guise, and credence* 1 in $\neg p$ relative to another. If we are to make sense of credences at all it seems we must consider them relative to a fixed guise. Unfortunately this doesn't seem to work either; relative to the guise corresponding to 'Hesperus is bright' I have a fixed credence in the proposition that Venus is bright, which is good, but no defined credence at all in the proposition that Mars is bright, or in other propositions which are not the content of a mental tokening of that kind of guise (presumably exactly those propositions distinct from the proposition that Venus is bright.)

1.1.5 Problems with hyperintensionalism

There are also specific objections that seem to apply only to the hyperintensional approach to propositions. David Lewis once objected to theories which tried to reduce modal facts to facts about language. Lewis's objection was aimed at linguistic theories of modality that identified possible worlds with certain kinds of sets of sentences. However, according to Lewis, there are simply more possible worlds than sets of sentences. A related point applies here. It certainly seems like there are more vague propositions than vague sentences. At the very least, there appear to be vague propositions which are not expressed by any sentence of English.

To see this imagine that we encounter an alien race who have evolved in a stellar system orbiting a white dwarf. Consequently the range of light that stimulates their visual system is in the ultraviolet spectrum. Like us they have different names for different segments of that region depending on how things appear to them. For example, just as we're unsure whether to apply the term 'red' on the region of the visual spectrum that is borderline red, there are corresponding regions of the ultraviolet spectrum where the aliens are unsure whether

to apply their terms. It seems very natural to say that there is something the aliens can say which we can't. These things aren't synonymous with the precise claims we can make about the ultraviolet spectrum in our language, and therefore encode pieces of information not expressible in English. If the vague propositions exist independently of us and the way we happen to speak, just as precise propositions do, the vague propositions we described would exist even if there weren't any aliens speaking this way. It seems particularly hard to imagine how the unknowability of these vague propositions could be explained by any sentence in English being borderline.

I think there is a much more pressing and general worry for the hyperintensionalist approach. According to hyperintensionalism there is a cluster of phenomena involving knowledge and other propositional attitudes that are associated with certain propositions and not others. The distinction is seen quite clearly in the case of the proposition that Harry is bald, which you cannot know, care about, rationally believe, and so on and the proposition that Harry has 2,000 hairs. For the sake of having a name for propositions like these, let us call these the 'vague propositions'. In a way this is an advantage of hyperintensionalism, as the intensionalist cannot even articulate this distinction.

If my arguments are correct, however, then whether a proposition is vague, i.e. whether it is can be associated with the cluster of phenomena mentioned above, does not hold in virtue of the relations that proposition stands in to sentences of a public language.

Thus for the linguistic theory which posits hyperintensional vague propositions there is both a distinction between vague propositions and vague sentences, neither of which are reducible to the other. Whatever we say about vague sentences does not allow us to eliminate the distinction between vague and precise propositions; a distinction which, if my arguments above work, is not reducible or explainable in terms of their relation to sentences.

However, if one accepts that there is a distinction between vague and precise propositions not explainable in virtue of the analysis of linguistic vagueness, then another simpler and more attractive analysis of linguistic borderline case becomes available.

> Sentence s is borderline as used in the language L by community C in context c at the the actual world at time t if and only if s expresses a borderline proposition in the language L, as used by community C in context c at the the actual world at time t

This reduction of the linguistic notion of borderlineness to the propositional notion seems to be a much better theory than one which posits two primitive kinds of borderliness, and then has to posit numerous principles connecting them. The above analysis eliminates the sentential notion, and the bridge principles are simply entailed by the definition. The linguistic notion seems redundant.

At any rate, whether we should accept the above reduction or not, we have at least motivated the primary goal of this thesis: that a theory of vague and precise propositions is not provided by linguistic theories yet is needed for an adequate theory of vagueness.

1.2 Outline of a non-linguistic theory

While most of the prominent theories of vagueness have been linguistic, there have been a few attempts to provide a non-linguistic theory of vagueness. For example Fine [43], Field [37], Barnett [6], Barnes and Williams [3], Kölbel [72] all treat indeterminacy, at various degrees of explicitness, as an operator expression, rather than a predicate applying to sentences. What is perhaps surprising is that a good proportion of these theories are primitivists about this operator: they generally think that there is very little to be said about its meaning other than, perhaps, some general principles about its logic and conceptual role.

For example, according to the theory defended in Barnes and Williams [3] the indeterminacy operator expresses a metaphysically primitive **operator** (just as predicates express properties, I'll use the bold faced '**operator**' for whatever operators express) in much the

same way as Prior thought that tense operators expressed metaphysically primitive **operators**. This leaves it somewhat mysterious as to why we must be ignorant in cases of indeterminacy and not, say, in cases of contingency. The account I am about to outline is one in which the **operator** simply couldn't be expressed by an operator with a different conceptual role. **Operators** and propositions according to this view are individuated by the role they play in a theory of rational propositional attitudes.

There has been a long standing dispute in the philosophy of language between truth conditional and conceptual role approaches to semantics. According to the former position the meaning of a sentence is given by the conditions under which it is true, a proposition. According to the latter position, the meaning of an expression or a thought is identified with the inferential relationship among rational thinkers/speakers it bears to other expressions and thoughts.

One of the most appealing aspects of truth conditional semantics, in the form of possible world semantics, is that it has enjoyed great success among linguists and is widely adopted in semantics. One of the characteristic features is that propositions are identified with sets of possible worlds. One of the drawbacks of these approaches, however, is that they have difficulty dealing with propositional attitudes. Conceptual role semantics on the other hand does a better job of accommodating propositional attitude reports, yet has proved to be unwieldy as a semantic theory.

I want to suggest a theory of meaning which draws something from both proposals.[14] I suggest that we retain the thought that the meaning of a sentence is given by a non-linguistic entity, a 'proposition.' However I reject the standard truth conditional account, which individuates these things modally. I think that propositions should rather be individuated by something analogous to a 'conceptual role', although, of course, something which applies to propositions and not sentences or thoughts. We can put the analogy as being between the role a proposition plays within a theory of rational propositional attitudes to the conceptual role a sentence or thought has in a conceptual or inferential scheme. The role of a proposition, p, is roughly analogous to the conceptual role of a belief, desire (or Ving) that p. This is partly given by the (objective) norms that governs one's propositional attitudes towards that proposition. Since the word 'conceptual role' has such strong linguistic connotations I shall henceforth use the word 'attitudinal role' for the role that a proposition plays in a theory of rational propositional attitudes.

An instance of this role is the evidential relationships a proposition has to other propositions. The attitudinal role of a proposition, in this sense, is not a linguistic role at all; the fact that, for example, the evidential probability that there are foxes given that there are vixens is 1, is one that applies to any rational being whether they speak a language (whether public or mental) or not.

Note that such a view does not preclude us from adopting the standard representation of propositions as sets of indices. Note, however, that we cannot think of the indices as maximally specific ways the world *could* have been. This assumption seems to load the dice in favour of natural language operators like 'could' and 'possibly', and it is no surprise that on this way of construing indices we run into trouble interpreting attitude operators like 'believes that' or 'desires that' in this framework. The assumption that indices are maximally specific ways the world could have been is not forced on us; we could just as easily interpret the indices as being maximally specific ways the world could be believed to be, or desired to be. One can still think of sets of such things as truth conditions: they are, after all, still propositions, i.e. *conditions* which can obtain or fail to obtain – they are the contents of belief, desire and so on, which can obtain of fail to obtain – and thereby secure or fail to secure the truth or falsity of the sentence they are assigned to. It is merely being denied that truth conditions are individuated by metaphysical possibility.

The success of the form of semantics which uses indices and relations to model operators therefore has nothing to do with the indices being interpreted as possible worlds. But the

[14] Although I do not think it is that important we think of it in these terms.

claim that the indices are ways the world could rationally be believed or desired (or more generally Ved) to be seems somewhat mysterious. A neutral, and less mysterious way to talk about the indices would be to say that they are 'maximally consistent' propositions: a proposition which is consistent and entails every consistent proposition that entails it. Here 'consistent' is not required to mean 'metaphysically possible' and 'entails', need not mean 'necessarily implies'.[15]

This takes away the air of mystery: on this way of talking indices are simply just certain kinds of propositions – maximally consistent propositions. And whether necessarily equivalent propositions are entailed by the same set of maximally consistent propositions depends once again on the question of how we individuate proposition – in this case, whether we individuate them as coarsely or more finely than necessary equivalence.[16]

On the proposed view the proposition that Harry is bald, call this p, is distinct from the proposition that Harry has less than N hairs, where 'N' is a precise name (e.g. a numeral) for the critical number of hairs it takes to be bald, call this latter proposition q. Under the simplifying assumption that baldness supervenes on hair number, p and q are necessarily equivalent, yet one may rationally have different credences in p and q. Indeed, I will argue in the proceeding chapters that according to any rational prior, the probability of p given Harry has exactly N hairs is strictly between 1 and 0 – any one who has evidence that Harry has exactly N hairs, has evidence that it's vague whether Harry is bald, and anyone who has evidence in this should be uncertain whether Harry is bald.

The proposition that Harry is bald and has exactly N hairs, and the proposition that Harry is not bald and has exactly N hairs, are both epistemically possible. Yet one is metaphysically impossible: either it is necessary that anyone with N hairs is bald, or it's necessary that anyone with N hairs is not bald (given simplifying supervenience assumptions.) Both of these two propositions are consistent, in the broad sense I have been elaborating on, only one is metaphysically possible (although it is vague matter which is the metaphysically possible proposition.) Since both these propositions are consistent, there are maximally consistent propositions i and j entailing the first and second of these propositions respectively; yet since one of these propositions is metaphysically impossible it follows that one of maximally consistent proposition i and j must be metaphysically impossible. In light of this, I shall call the indices, like i and j, (im)possible worlds, to indicate that they are maximally consistent propositions that may or may not be metaphysically impossible.

There should be nothing metaphysically mysterious about impossible worlds in this sense; anyone who thinks there is more than one impossible proposition and also thinks there are maximally strong propositions that are not inconsistent can make sense of this notion.

It's worth noting that even without bringing in considerations to do with attitudes the style of semantics which employs indices and accessibility relations may not always allow one to interpret the indices as possible worlds. For example, a purely modal language can only allow the indices to be interpreted as representing possible worlds if we are assuming a modal logic including the $\mathsf{S4}$ principle. If we are taking the index semantics at face value then only the indices accessible to the index representing the actual world will be genuine 'possible' worlds. Similar points hold for counterfactual logics which allow for non-trivial counterfactuals with impossible antecedents. Neither of these two examples rely on the features of propositional attitudes.

[15] Of course, if necessarily equivalent proposition are identical, then by Leibniz's law p is consistent iff q is consistent whenever p and q are necessarily equivalent. A notion of consistency which forces propositions to be more fine grained than sets of possible worlds is just not coherent unless one already takes propositions to be more fine grained than sets of possible worlds.

[16] Note that given certain idealisations the notion of propositional identity and the notion of consistency appealed to here are interdefineable. For example we could define the broadest notion of necessity, Lp as $p \equiv \top$, and the broadest notion of possibility Mp as $\neg M \neg p$. (Note we also get $\mathsf{S4}$ assuming $(p \equiv p) \equiv \top$: from $p \equiv \top$ and Leibniz's law we get $(p \equiv \top) \equiv \top$, i.e. $Lp \rightarrow LLp$.)

1.2.1 The theory of propositions

The theory of propositions and (im)possible worlds follows from a broadly Stalnakerian way of making sense of the indices.[17] According to the Stalnakerian view [128] we begin with a theory of rational agents in which indices are taken as primitive. Whatever structure indices have is abstracted from this theory, and beyond this their nature is left open.

Stalnaker is not explicit about what this theory is exactly, but he does say the following:

> "What is essential to rational action is that the agent be confronted, or conceive of himself as confronted, with a range of alternative possible outcomes of some alternative possible actions. The agent has attitudes, pro and con, towards the different possible outcomes, and beliefs about the contribution which the alternative actions would make to determining the outcome. One explains why an agent tends to act in the way he does in terms of such beliefs and attitudes. And, according to this picture, our conception of belief and of attitudes pro and con are conceptions of states which explain why a rational agent does what he does." Stalnaker, Inquiry, p5. [128]

On this picture of the metaphysics of indices, they are not things which are deeply tied to the world, as a possible world in Lewis's sense would be: "they obviously are not concrete objects or situations, but abstract objects whose existence is inferred or abstracted from the activities of rational agents" ([128] p50-51.)

From the above passage you would be forgiven in thinking that Stalnaker individuates the 'possible outcomes' according to a rational agents bouletic attitudes ('attitudes, pro and con') and doxastic attitudes. It is worth noting, however, that Stalnaker individuates propositions, and possible outcomes, modally, and indices in the sense abstracted above are identified with possible worlds. Stalnaker's reasons for doing this seem to have nothing to do with the picture outlined so far, and have more to do with his reductionist tendencies with respect to the problem of intentionality. He even considers the impossible worlds approach, writing 'could we escape the problem of equivalence by individuating propositions, not by genuine possibilities, but by epistemic possibilities – what the agent takes to be possible?' but dismisses it on the grounds that although 'this would avoid imposing implausible identity conditions on propositions [...] it would also introduce intentional notions into the explanation, compromising the strategy for solving the problem of intentionality.'[18]

However, it seems obvious to me that the entities we abstract from a theory of rational decision will be more fine grained than possible worlds.[19] An astronomer who believes that Hesperus is a planet may display very different behaviour, both linguistically and non-linguistically, from someone who believes that Phosphorus is a planet. It furthermore seems perfectly possible that this astronomer may have very good evidence that Phosphorus is a planet, which is not also evidence that Hesperus is. In which case it seems to me to be perfectly rational for this agent to believe that Hesperus is a planet without believing that Phosphorus is and rational for her to act accordingly.

Since I am claiming that propositions are more fine grained than possible worlds it's natural to ask at what point do we stop individuating? How fine grained are propositions? I think that Stalnaker's theory provides a perfectly principled answer to this question: indices are as fine grained as we need them to be to specify the functional role of the belief that p in terms of the indices at which p is true. In other words, indices are whatever they need to be so that functional roles do not distinguish more finely than sets of indices.

[17] Of course, for Stalnaker indices are metaphysically possible worlds.

[18] Note: in the more recent [127] he disavows the project of reduction.

[19] Stanley [129], makes the similar point that Stalnaker's 'causal pragmatic' approach to intentionality appears to be compatible with, and even motivates, a theory in which indices are epistemically possible worlds.

The functional role of a belief that p does not include the role that a belief that p plays in the cognitive workings of completely erratic and irrational agent (if we could ever determinately attribute a belief that p to such an agent.) No two belief states have the same cognitive role across every possible agent, rational and irrational; the resulting notion of proposition would be uninteresting and useless for the purposes of giving a theory of objective information that can be communicated among different agents. For example, in order have a theory of communication some degree of rationality must be assumed by both participants in a conversation if we are to be able to infer anything about what they believe from the sentences they are uttering.

The words 'rational', 'should', 'ought', and so on, which we use in everyday life, are vague and highly context sensitive. I believe this context sensitivity persists even if we prime our audience to think we are using 'ought', say, to mean something broadly epistemic, as opposed to moral; e.g. 'she ought to raise as she knows he is bluffing', 'she ought to fold; her opponent has lost most of his money already'. Therefore an important component of the theory of propositions I am proposing is a theory of the kind of rationality at play. In what follows I shall try to give a formal and an informal account of 'coherence' – having one's prior beliefs and desires compatible with the rational constraints I am interested in – which I think we already have some intuitive grasp over. However, I am not proposing to explain this notion in more basic terms: the notion of coherence I shall outline is a primitive of my theory from which other concepts (being a precise proposition, et cetera) will ultimately be explained. It is only the notion we express with the words 'rational' or 'should' in special contexts. It should be no surprise that this proposal for individuating propositions has primitives. Even the modal way of individuating propositions must either take metaphysical modality as primitive, or introduce further notions to analyse this kind of possibility.

The other function of this description is to provide the basis of a decision theory from which indices and propositions are to be abstracted. That is: I shall outline a decision theory in which I freely quantify over propositions and indices, with the aim of at least implicitly defining these notions.

The primitives of our theory thus consist of[20]:

1. A set of indices, $\mathcal{I}$.
2. A set of regular probability measures, Coh over the powerset algebra $\mathcal{P}(\mathcal{I})$, called the *coherent priors*
3. A set of real valued functions $u : \mathcal{I} \to \mathbb{R}$, U, called the *coherent utilities*.

We begin by making two definitions:

Definition 1.2.0.1. *Let Pr be a coherent prior. The evidential probability for an agent at a time t relative to Pr is $Pr(\cdot \mid e_t)$ where e_t is the conjunction of the agents evidence at t.*

Definition 1.2.0.2. *Let Pr be a coherent prior and u a coherent utility. The initial preference ranking given by Pr and u is given by the propositional relation: $p \prec q$ iff $E(u \mid p) < E(u \mid q)$. $E(u \mid p)$ stands for the conditional expectation of the random variable, u, given p.*

For any proposition, e, we define the agents conditional preferences on e: $p \prec_e q$ as $p \wedge e \prec q \wedge e$.

An agent is rational only if there exists a coherent prior, Pr, and coherent utility function u such that:

(a) The agents credences at time t are identical to her evidential probabilities at t relative to Pr; i.e. $Pr(\cdot \mid e_t)$ where e_t is the conjunction of all her evidence at t.

[20] I am, for the moment, ignoring the complexities that arise when the set of indices infinite.

(b) The agents preferences at time t are given by $p \prec_{e_t} q$ (i.e. by the relation $E(u \mid p \wedge e_t) < E(u \mid q \wedge e_t)$.)

The formal constraints we have outlined already tell us something informative about coherence and about rationality. For example, every rational agent must have credences obtained by conditioning on coherent priors; from this it follows that the credences of a rational agent are probability functions. Among other things it follows that every rational agent has credence 1 in every tautology.

I suggest further that every rational agent has credence 1 in propositions expressed by paradigm examples of 'analytic' sentences. Therefore, every coherent prior:

1. Assigns 1 to the proposition that all vixens are female foxes
2. Assigns a number to the proposition that there are vixens which is less than or equal to the proposition that there are foxes.
3. Assigns 1 to the proposition that Harry is bald conditional on the proposition that Harry has 0 hairs.
4. Assigns a number strictly between 0 and 1 to the proposition that Harry is bald conditional on Harry having N hairs.

Facts like the above are not entailed by the formal constraints outlined above, but are just as important to the notion of coherence I am characterising: roughly that of being *conceptually* coherent. A coherent prior is thus a probability function which is both conceptually coherent, and a regular prior (i.e. does not assign a probability of 0 to any (im)possible world.) It follows from the first claim, for example, that the evidential probability that all vixens are female foxes, for any conceptually coherent agent at any time, is 1. Of particular interest is the claim, 4., that no coherent prior can assign the proposition that Harry is bald a credence of 1 or 0, conditional on the proposition that he has N hairs, where N is suitably chosen so that the proposition that Harry has N hairs conceptually entails that it's vague whether Harry is bald. I take it that this kind of constraint is of a similar kind to claim 3. – that no coherent prior may assign a number less than 1 to the proposition that Harry is a fox conditional on the proposition that Harry is a vixen. Roughly speaking, a probability function is coherent if it respects all conceptual requirements of the kind including 4., of which conceptual entailments are just a special case.

1.2.2 Precision operators and determinacy operators

The aim of this thesis is to elucidate, within this framework, the notion of a proposition (i.e. a set of (im)possible worlds) being precise.

It is worth spending some time clarifying what I mean by a precise and a vague proposition. Many philosophers theorise instead with the notion of a propositions being determinately true, which they represent with an operator Δ. To say that p is determinate is roughly to say both that p and p is not borderline. To see the difference consider the proposition that Henry is bald. Henry, unlike Harry has absolutely no hair at all. The proposition that Henry is bald is therefore determinately true, it is true and it is not borderline. Nonetheless there is a clear and intuitive sense in which the proposition that Henry is bald is a vague proposition – after all Henry could have been more like Harry actually is, and had instead a borderline number of hairs. Neither the proposition that Harry is bald or the proposition that Henry is bald is a precise proposition.

In contrast, the current approach takes the notion of a precise proposition as primitive. According to this approach we may characterise formally a notion of precision from which the Δ operator can be defined. For example, it is clear that the negation of a precise proposition is precise, and a conjunction or disjunction of precise propositions is precise. Thus the

precise propositions form a complete Boolean subalgebra of the algebra of propositions.[21] Principles like these allow us to rigorously formulate a logic of precision analogous to the more traditional logics of determinacy. There are a number of advantages to this approach which I shall outline below.

For the kind of theorist I have described above it is often assumed that the notion of a precise proposition is recoverable once we have the resources of an operator representing metaphysical necessity, $\Box$. The thought is that while in the example above the proposition that Harry is bald and the proposition that Henry is bald, call these p and q, differ with regard to determinate truth, they are both possibly borderline. Perhaps this is what it takes for a proposition to be vague, and that a proposition is precise iff it's not possible that it is borderline. Let $\sharp p$ (pronounce: 'sharp p') be an object language operator stating that p is precise. In this case the notion of a proposition being precise might be defined as $\sharp p := \Box(\Delta p \vee \Delta\neg p)$.

This definition is formally adequate: arbitrary conjunctions, disjunctions and negations of necessarily determinately true or determinately false propositions are necessarily determinately true or determinately false (assuming a reasonable logic of determinacy and necessity.) However it is inadequate to capture the intuitive notion we began with. Consider a variant of the example involving Henry: let p be the proposition that water is has chemical composition XYZ and Henry is bald. This proposition is determinately false, but unlike the original example, this feature is also necessary. However, the following seems like a completely coherent scenario: water might have turned out to have a different chemical constitution than it in fact has, XYZ, and Henry might have turned out to have a number of hairs that is borderline. Therefore, I suggest, the proposition p is also vague despite being necessarily determinately true or determinately false.

The failure of this definition to capture precision, in my opinion, should make one doubtful that the notion can be defined from the determinacy operator at all. On the other hand, given the notion of a precise proposition one can define what it means for a proposition to be determinately true:

> It's determinate that p if and only if the strongest true precise proposition entails p.

thinking of p as just a set of (im)possible worlds the notion of entailment is simply that of one proposition being a subset of the other, and the notion of strength being that determined by entailment.

Let me end by distinguishing my notion of precision from another which has been proposed in Williamson [142]. The claim that p is precise, on my view, entails that it is necessarily determinately true or determinately false, i.e. $\Box(\Delta p \vee \Delta\neg p)$, even though I have argued that the converse doesn't hold. The concept of precision defined in [142], however, is much stronger: p's being precise, according to that definition, would entail that every instance of the schema $\bar{O}(\Delta p \vee \Delta\neg p)$ held, where $\bar{O}$ can be substituted for an *arbitrary* sequence of operators consisting of Δ's and $\Box$'s. Call this notion 'uber-precision'. If N is a clearly small number which is not clearly, clearly, clearly (and so on 100 times) small, then a putative example of a precise proposition which is not uber-precise is the proposition that N is small. You might argue that N is precise because if N is clearly small, it's necessarily clearly small (numbers have their 'size' necessarily). Although this argument relies on the inadequate definition of precision in terms of $\Box$ and Δ it is very suggestive. Of course, given the inadequacy of the former definition of precision in terms of $\Box$ and Δ, the latter definition of uber-precision in terms of $\Box$ and Δ should also be suspect. Another way to distinguish the two notions is as follows. Let B_n be the proposition that having less than n hairs is sufficient for being bald. Then precision lines up with the notion of determinate truth in the following way: the last n such that B_n is determinately true is identical to the last n such that B_n is precise. The last n such than B_n is uber-precise is usually much

[21] Although, since it may be vague whether a proposition is precise or not, propositions about which propositions are elements of this algebra will not always be precise.

smaller, and would be equal to the last n such that B_n is determinately true at all orders (i.e.: determinately true, determinately determinately true, and so on.)

1.2.3 Logical issues

Here is the theory of precision, CID, for the language of propositional logic (henceforth: PC) augmented with the $\sharp$ operator (PC$\sharp$.)

RE If $\vdash \phi \leftrightarrow \psi$ then $\vdash \sharp\phi \leftrightarrow \sharp\psi$

RN If $\vdash \phi$ then $\vdash \sharp\phi$

C $\sharp\phi \wedge \sharp\psi \rightarrow \sharp(\phi \wedge \psi)$

D $\sharp\phi \wedge \sharp\psi \rightarrow \sharp(\phi \vee \psi)$

I $\sharp\phi \leftrightarrow \sharp\neg\phi$

The intended models of CID are models of the form $\langle \mathcal{I}, e, [\![\cdot]\!] \rangle$ where $\mathcal{I}$ is a set of (im)possible worlds, e is a function from $\mathcal{I}$ to equivalence relations over $\mathcal{I}$ and $[\![\cdot]\!]$ a function from propositional letters to subset of $\mathcal{I}$. We define $[\![\sharp\phi]\!] := \{i \in \mathcal{I} \mid \forall j, k(j \in [\![\phi]\!], \langle j, k\rangle \in e(i) \rightarrow k \in [\![\phi]\!])\}$. The semantic values of disjunctions, conjunctions and negations are calculated as usual. Intuitively the partition $e(i)$ is the set of maximally strong precise propositions according to i, and a proposition is precise at an index i iff it is a union of maximally stronge precise propositions. CID is sound and complete relative to this class of models.

If we add a notion of entailment, $\models$ and propositional quantification to the language (call this language: PC$\forall\sharp\models$) we may define the determinacy operator as follows

$$\Delta p := \forall q((\sharp q \wedge q \wedge \forall r(\sharp r \wedge r \rightarrow q \models r)) \rightarrow q \models p).$$

Note that since propositions stating that a proposition is precise, need not themselves be precise, the Δ operator defined above need not iterate trivially. In order to prove that Δ obeys the modal logic KT (see below), we also need to add the principle stating that there always is a strongest true precise proposition, and some basic facts about entailment:

$\exists p(p \wedge \sharp p \wedge \forall q(q \wedge \sharp q \rightarrow p \models q))$

$p \wedge p \models q \rightarrow q$

$p \models q \wedge p \models r \rightarrow p \models q \wedge r$

$p \models q \leftrightarrow p \models \neg\neg q$

For logical purposes we can take Δ as primitive (even though it is not philosophically basic) and theorise about its logic alone. Let PCΔ be the language of propositional logic augmented with Δ in the usual way. The logic of this language I shall assume through out the thesis is KT, which, in addition to being closed under classical propositional logic and uniform substitution, contains the principles:

RN If $\vdash \phi$ then $\vdash \Delta\phi$

K $\Delta(\phi \rightarrow \psi) \rightarrow (\Delta\phi \rightarrow \Delta\psi)$

T $\Delta\phi \rightarrow \phi$

In the pure language of the propositional calculus with the Δ operator the theory I defend does not look that much different from that of a supervaluationist logic. Throughout this thesis it will become clearer how my view differs from supervaluationism. However, even the logical properties of my theory of determinacy may differ from standard supervaluationism once we add either

- Propositional quantifiers.
- Modal operators.

I shall consider these in turn.

The theory of propositional quantification can be stated by adding to the propositional calculus, PC, an infinite stock of propositional variables and a propositional quantifier $\forall$ binding them, PC$\forall$. Add to classical proposition logic:

PGen If $\vdash \phi$ then $\vdash \forall p\phi$.

PUI $\forall p\phi \rightarrow \phi[\psi/p]$

PK $\forall p(\phi \rightarrow \psi) \rightarrow (\forall p\phi \rightarrow \forall p\psi)$

PR $\phi \rightarrow \forall p\phi$ provided p does not occur unbound in ϕ.

In general you may have a modal logic in the language PCO that is complete relative to possible worlds style semantics, and also have a similar possible world semantics for the logic of propositionally quantified logic in the language PC$\forall$. However, when you add the two theories together they will typically not be complete for a possible worlds semantics; there are principles which involve the interaction of the modalities and quantifiers which are validated but are no longer provable. For example, in a propositionally quantified modal logic of S5 you cannot prove that necessarily there is a proposition which entails all truths (i.e. that there are maximally strong propositions, i.e. atoms or 'possible worlds') yet this is validated in every possible worlds model. The same applies here; we cannot prove that there is a proposition which determinately implies all truths in the joint language PC$\Delta\forall$. We must also add the principle:

At $\exists p(p \wedge \forall q(q \rightarrow \Delta(p \rightarrow q)))$

To see how the theory might diverge from a standard supervaluationist framework lets consider the consequences of the axiom PUI. Defining $\exists p$ as $\neg\forall p\neg$ we can prove a propositional comprehension principle:

PComp $\exists p\Delta(p \leftrightarrow \phi)$

PComp^n $\exists p\Delta^n(p \leftrightarrow \phi)$

Here Δ^n is notation for a sequence of n of occurrences of Δ. Note firstly that $\forall p\neg\Delta^n(p \leftrightarrow \phi) \rightarrow \neg\Delta^n(\phi \leftrightarrow \phi)$ by PUI, and secondly that $\Delta^n(\phi \leftrightarrow \phi)$ by n applications of RN. Thus we may infer $\neg\forall p\neg\Delta^n(p \leftrightarrow \phi)$ by modus tollens.

PComp entails that there is some proposition determinately equivalent to the proposition that Harry is bald. And PUI entails the existence of vague propositions: for example if it's vague whether Harry is bald then PUI entails there is a proposition which is vague ($\nabla\phi \rightarrow \exists p\nabla p$.) These are controversial principles, and certain versions of supervaluationism might want to deny them. According to Fine [43], the proposition that Harry is bald is neither true nor false. One might therefore think there is simply no fact, i.e. no such thing as the true proposition, that Harry is bald or that Harry is not bald. Such theorists may instead want to replace PUI with

PRUI $\forall p\phi \rightarrow \phi[\psi/p]$ when ϕ does not contain Δ.

In Fine [42] this this is called 'restricted specification' (see that paper for further details and semantics.[22])

[22]The semantics which invalidate PUI allows the interpretation of propositional constants and propositional variables to belong to and vary over a Boolean algebra of sets of the set of worlds $\mathcal{I}$. PUI may be invalidated when this algebra is strictly small than the powerset algebra.

Let us now consider the result of adding operators representing both metaphysical necessity and determinacy to the propositional calculus, PC$\Box\Delta$. Here, once again, we encounter differences in motivation between supervaluationist semantics and my own – this time due to the interaction between vagueness and modality.

On the standard supervaluationist semantics the indices are represented as pairs of worlds and precisifications. Precisifications are interpretations of the language in question telling us what the extensions of vague predicates are at different worlds; they are tied to a particular language, such as English, and there is no straightforward connection between the existence of a world precisification pair in which p holds, and the existence of an epistemic possibility in which p holds.

For the supervaluationist the indices consist of two separate domains: the set of possible worlds, W, and a set of precisifications, V, with the indices $\mathcal{I}$ being identified with $W \times V$, the set of all ordered pairs whose first element is in W and whose second element is in V. W will usually be thought to come with the universal accessibility relation, E, (more generally, E could be an equivalence relation) and V with a reflexive relation of relative admissibility: $R \subseteq V \times V$. Ruv means that if we interpret 'admissible' according to u, v would be admissible.

In order to model both modality and vagueness simultaneously we get a derived model, $\langle \mathcal{I}, E', R' \rangle$ defined as

$R'\langle x, u\rangle\langle y, v\rangle$ iff $x = y$ and Ruv

$E'\langle x, u\rangle\langle y, v\rangle$ iff Exy and $u = v$

R' and E' can be used to model the operators Δ and $\Box$ respectively in the richer language of PC$\Box\Delta$. Note that even if E is the universal relation on W, E' is a non-universal equivalence relation whenever $|V| > 1$.

Frames such as the one described above are known as 'product frames'. By some standard results it follows that the logic of frames generated like this from arbitrary equivalence relations, E, and reflexive relations, R is

- KT for Δ
- S5 for $\Box$

ND $\Box\Delta\phi \rightarrow \Delta\Box\phi$

DN $\Delta\Box\phi \rightarrow \Box\Delta\phi$

PD $\Diamond\Delta\phi \rightarrow \Delta\Diamond\phi$

Note that the dual of the final axiom, $\neg\Delta\neg\Box p \rightarrow \Box\neg\Delta\neg p$, follows from the final axiom by contraposition. Note that the converse of the final axiom is clearly not valid. Suppose there is a bag of balls numbered from 1-1000. Determinately, for any n it's possible that I have picked the nth ball. Since some n is the last small number it follows that, determinately, it's possible that I've picked the last small number. However it is not possible that I've determinately picked the last small number. Note that the dual of the converse of the final axiom entails the converse of the final axiom (by contraposition) so the dual of the converse of the final axiom is also false. The above result would also hold if we either substituted KT for KTB, S5 for S4, or both, and accordingly substituted 'reflexive' frames for reflexive symmetric frames frames, and 'equivalence relation' for 'transitive reflexive'.[23]

On the current view models are not generated like this. The indices, I, are simply treated primitively as (im)possible worlds with a equivalence relation S and a reflexive relation R. Since there are impossible worlds, and possible worlds, S will be non-universal.

[23]See Gabbay and Shehtman [47], and Kurucz [77]. By the results of Gabbay and Shehtman, we can substitute any PTC logic S5 and KT, provided we substitute the models with the classes they validate.

$\Box\Delta\phi \rightarrow \Delta\Box\phi$	$S \circ R \subseteq R \circ S$
$\Delta\Box\phi \rightarrow \Box\Delta\phi$	$R \circ S \subseteq S \circ R$
$\Diamond\Delta\phi \rightarrow \Delta\Diamond\phi$	Church-Rosser Property

Table 1.1: Vagueness and modality

In this setting the principles of supervaluationist modal logic do not simply fall out of the structure of indices, but are substantive constraints on the accessibility relations. These constraints are listed in table 1.2.3 – the Church-Rosser property refers to the condition: $(Rxy \wedge Sxz) \rightarrow \exists w(Syw \wedge Rzw)$ and $R \circ S$ denotes the composition of R with S, $(R \circ S)xy \leftrightarrow \exists z(Rxz \wedge Szy)$.

The logic of modality and vagueness I suggest as a starting point may simply be $\mathsf{S5}\Box + \mathsf{KT}\Delta$.

Note that for the supervaluationist there is a sharp distinction between vagueness related facts and worldly facts. Suppose that $\mathcal{I} = W \times V$. Let $x \in W$ be a possible world. One might think that a proposition, i.e. a subset of $\mathcal{I}$, denotes a possible world iff it is of the form $\{\langle x, v\rangle \mid v \in V\}$. Note that if p designates a class of this kind in a model then $p \rightarrow \Delta p$ and $\neg p \rightarrow \Delta\neg p$ will hold in the model; it is always a determinate matter what a possible world is in this sense (this idea, that there are possible worlds like this, leads to problems to do with higher order vagueness.)

We can define the notion of a possible world in a neutral way in both settings in the language $\mathsf{PC}\forall\Box\Delta$, by adding propositional quantification to the language: a possible world is a possible proposition which strictly implies every possible proposition or its negation. This can be stated in the object language if we include propositional quantifiers. It's natural to think that p is a possible world just in case p is possible and it strictly implies every proposition that strictly implies p: $\Diamond p \wedge \forall q(\Box(p \rightarrow q)) \rightarrow \Box(q \rightarrow p)$. Actually, since there can be distinct propositions which strictly imply each other, it is more natural to define a possible world to be not only a proposition which strictly implies every proposition that implies it, but to be the the weakest such proposition. Note that in both settings, which propositions are possible worlds is a vague matter in the sense that we do not get $\Delta PW(p) \vee \Delta\neg PW(p)$. In the supervaluationist setting, therefore, this notion of possible world doesn't line up with the notion defined above. For example: if N is the last small number, then N is small at every possible world, since facts about whether numbers are small are not contingent. Yet since it's vague whether N is small it's vague whether it's possible that N is small, and so for every possible world that entails that N is small (all of them) it's vague whether that proposition is possible, and thus vague whether it's a possible world.

This seems to be a puzzling feature of the supervaluationist approach: the formalism of making models by taking pairs of worlds and precisifications makes it seem like the set W consists of real possible worlds which can be sharply distinguished from the ways in which a proposition can vary its truth value due to vagueness – the set of precisifications. However, what could a possible world be if not a possible proposition that is maximally strong in the order of strict implication (or, at least, something closely related to this kind of proposition)? Supervaluationism is commited to a precise distinction between the worldly facts and the vagueness related facts, when no precise distinction like this exists.

In an arbitrary (im)possible worlds frame there is simply no way of getting back to a notion of possible world such that it's a precise matter which propositions the possible worlds are. Is it a problem that we cannot sharply distinguish possible from impossible worlds? A picture in which this might be problematic is if we had this conception of worlds as concrete physically existing objects, and epistemically possible worlds as some how derivative constructions reducible to these concrete things (e.g. centered worlds, pairs of worlds and precisifications, etc.) Vagueness over what possible worlds there are would be vagueness concerning what things there are simpliciter, which seems to be strange kind

of ontic indeterminacy.

The alternative view takes worlds simply to be certain kinds of propositions. This avoids modal realism. However, given that propositions are individuated by propositional attitudes, among other things, we end up with more these maximal proposition like things hanging around than there are possible worlds. Possibility is simply a predicate which applies to some of these proposition and not others – in this setting vagueness about what is possible does not involve vagueness about what exists.

Chapter 2

Vagueness and Evidence

In the last chapter we argued against linguistic accounts of vagueness on the grounds that if it is true that we cannot know whether Harry is bald, then this is the sort of thing that would be true no matter how we happened to use the sentence 'Harry is bald'. Whether a belief that Harry is bald constitutes knowledge or not, I suggested, is just not sensitive to facts about your linguistic environment.

Whether or not vagueness is linguistic, there is a quite general puzzle as to how we acquire vague beliefs at all. It is important to get clear on what role language plays in the acquisition of a *belief* that Harry is bald if we are to assess the various merits and demerits of the two approaches to vagueness. Speaking on their behalf, how might a linguistic theorist respond to this challenge? A natural suggestion, one that I have heard in conversation, if not in print, goes something like this: in order to have a belief that Harry is bald you need to be able to internally token a certain kind of sentence in your mental language; one that is synonymous with, or has the same conceptual role as the English sentence 'Harry is bald'. In order for a mental representation to acquire this kind of role it has to stand in the right kind of relation to a vague public language sentence.

This explanation would evidently rule out a non-linguistic theory such as my own. Whether one has vague beliefs at all, according to this proposal, is dependent on one speaking a public language.[1] But the proposal is controversial. Burge has argued [16] that whether one has beliefs involving arthritis and other 'deferential' concepts can be sensitive to your linguistic environment. While these arguments have enjoyed a moderate amount of acceptance among philosophers, an argument that purports to show that vague beliefs are also like this would threaten to show that we could have barely any beliefs at all unless we spoke a public language.

There are plenty of other things to say about this proposal. My purpose in this chapter is rather to highlight other mechanisms by which we acquire vague beliefs which the linguistic theorist cannot account for. I shall argue that most of our beliefs do not require any familiarity with a public language; most of our vague beliefs are acquired via our sensory faculties, vision, smell, proprioception, or other non-linguistic faculties such as memory. Sometimes public language sentences express vague propositions and these sentences can be used to communicate vague beliefs, but the class of vague propositions which satisfy the evidential profiles outlined vastly outstrips the class of propositions expressed in any given language and, I shall argue, and the way in which language features in our acquisition of vague beliefs is minimal.

In section 2.1 I introduce the idea of one's evidence being 'inexact' and argue that there are important connections between these cases and cases in which one's evidence consists

[1] It should be noted that it is unclear whether this suggestion meets my original objection (that whether a true belief that Harry is bald could constitute knowledge is independent of your linguistic environment.) Maintaining that one cannot believe that Harry is bald unless one speaks a public language does nothing to further clarify when such beliefs count as knowledge or not.

of vague propositions. In section 2.2 we discuss Jeffrey's alternative to the conditioning update rule and argue that conditioning on a vague proposition is equivalent to Jeffrey conditioning on a set of precise propositions (its 'precisifications'.) In 2.3 I argue that one's beliefs and credences in vague propositions are, in a certain sense, determined by your beliefs and credences in the precise. While in 2.3 I argue that each vague proposition has a certain evidential role, determined by they way in which credences in that proposition depend on your credences in the precise propositions, in section 2.4 I propose an existence principle for propositions guaranteeing that for any possible evidential role there is some vague proposition which has that role. Finally I turn to some miscellaneous objections and questions about the view in 2.5.

2.1 Inexact evidence

Suppose you are looking out of your window at a tree in the distance. In doing so you obtain some knowledge about the tree. However, your knowledge is not exact; for example, while you now know that the tree is larger than 10cm and less than 1000cm, based on what you can see the exact height of the tree in cm is still unknown to you.

In [137] Timothy Williamson introduced the term 'inexact knowledge' for the kind of epistemic state you would be in in situations like the one above. A parallel distinction arises with respect to your evidence in such situations. Given that you have only seen the tree from a distance, and your eyesight is not perfect, your evidence includes some propositions – that there is a tree, that it's larger than 10cm and smaller than 1000cm – but does not include the proposition stating the trees exact height. We may call your evidence in such cases inexact evidence.

Being inexact is not the same as being inaccurate; if the tree appeared to be less than 500cm when it was in fact greater than 600cm then my evidence, or my apparent evidence, would be inaccurate, but in the case described in the opening paragraph my beliefs are not inaccurate. All of my evidence, or apparent evidence, is accurate. On the other hand, if I went out and measured the tree, the relevant source of inexactness in my evidence about the tree's height would eliminated. But the phenomenon at hand is not ignorance or lack of evidence about the tree's exact height. If I had never seen the tree, but had been informed by a highly reliable person that the tree was less than 600cm then my evidence would be exact, even though I still do not know the height of the tree. Nor is the relevant sense of exactness specificity: if I have been told that the tree is between 400cm and 500cm then my evidence is more specific than it would have been if I had only been told that the tree was between 300cm and 600cm, but the latter case is not closer in kind to the state of evidence you would find in the opening example.

I hope it is clear that the evidential status of someone who has seen a tree from a distance is relevantly different from any of the preceding examples. While I cannot give an uncontroversial definition of what must be involved in cases where one's evidence is inexact, I hope the preceding examples have elucidated it enough to provide a good enough handle on the kind of situation I am interested in to be worth discussing further.

Evidence obtained via imperfect sensory faculties are paradigm examples of inexact evidence. However, it is not clear that it is always the case that when your evidence is inexact, some sensory faculty is at fault. Having taken a cursory look at my bookcase, I gain inexact evidence about the number of books I own; I have a general feel for how many books there are but I do not know exactly how many books there are without having counted. However it is not completely obvious that my evidence would be significantly better if I had perfect vision. It seems more natural to say that it was not my visual experience, but rather my ability to process it, which was to blame.

I take it that most ways of obtaining evidence leave room for the possibility that evidence

obtained in that way is inexact.[2] I also take it that most of our evidence *is* inexact. The main thesis of this chapter is the claim that inexact evidence is vague evidence; that when we find ourselves with inexact evidence the strongest proposition we ought to be certain of on the basis of this evidence is a vague proposition.[3] This partly vindicates the claim that we do not need to be acquainted with a public language to have vague beliefs. Furthermore, since almost all of our evidence is inexact, it follows that vague propositions occupy a very important evidential role. They are usually the best pieces of information we have when we learn from experience, they are what we usually reason from and to, and they are the objects of our desires upon which we act. If it is the evidential role of a proposition that is partly constitutive of what it is to be that proposition, then our thesis also gives us some insight into what vagueness and vague propositions are.

The claim that inexact evidence is vague evidence requires some refining however. If I have examined a man's head carefully and have determined that he has no hairs at all, then I have evidence that he is bald. In such a scenario my evidence about the man's hair number is exact, yet my evidence includes the vague proposition that the man is bald. Having evidence for a vague proposition is thus not sufficient for being in the kind of circumstance characteristic of inexact evidence. However, the proposition that the man is bald is not the strongest proposition I have evidence for – the strongest proposition is the proposition that the man has no hairs. While this proposition entails that the man is bald, it is not equivalent to it: a man with one short hair is bald, so one can be bald without having no hairs at all. The claim I am interested in is rather the claim:

> When your evidence about a subject matter is inexact, the strongest proposition about that subject matter that your evidence fully supports is a vague proposition.

One might object that in the cases described, my evidence is not propositional at all, but rather consists of an imprecise visual experience, a hazy memory, or some other non-propositional object. As the claim is stated, however, I have not assumed that all evidence is propositional. Even if your evidence consists in visual experiences, memories, or what have you, what that evidence supports, I assume, is always a proposition. We may ask what the strongest proposition supported or justified by that evidence is, without assuming that the proposition in question is your evidence. In what follows, however, I shall often simply talk about an agents evidence as if it were just a set of propositions – everything I say can be rephrased to be said about the set of propositions my evidence justifies.

Let us concentrate on some specific ways of obtaining evidence.

1. S saw that p
2. S remembered that p
3. S could hear that p
4. S could feel that p
5. S could see that p
6. S found out that p

These are all paradigm cases of what I shall call 'evidential attitudes': an attitude one cannot have towards p without thereby having evidence that p.[4] In cases 3., 4. and 5. the auxiliary 'could' usually marks that the verb is being understood perceptually. Furthermore, evidential attitudes are characteristically factive, whereas 'heard that' and

[2] One might argue that certain *a priori* methods, such as mathematical proof, never leave one with inexact evidence. I would dispute this, but we can set this case aside for the purposes of this chapter.

[3] I shall say that a proposition, p, as vague iff it is possible that it's vague whether p, and that it's precise otherwise.

[4] Here I borrow heavily from the discussion of 'factive mental state operators' in Williamson's [143].

'felt that', without the modal, are not factive or evidential. 'Jane heard that Hector is getting fired, although she didn't have any reason to believe that he was getting fired' seems to be a fine thing to say if, for example, Jane's source was known to be completely unreliable, whereas 'Jane could hear that Hector was getting fired, although she didn't have any reason to believe that he was getting fired' seems to be harder to maintain. The latter sentence implies that the actual firing is directly audible to Jane, whereas the former sentence implies that she has heard second hand that Hector is being fired, but hearing in this way may or may not have any evidential import. On the other hand, both 1. and 5. are factive and evidential. The auxiliary in 5. can usually be inserted to force the reading where the evidence at hand is perceptual. This is not always the case, however, as seen by the sentences 'Hector saw that the theorem was true' and 'Hector could see that the theorem was true' – both seem to be ok things to say, yet in neither case is the kind of seeing a perceptual one. The fact remains, however, that whether perceptual or not, one cannot see that p without p becoming part of your evidence.

It should be pointed out that one does not need to be particularly linguistically competent to have any of these attitudes towards a proposition. One could hear (see, remember, feel) that p without speaking any particular public language or standing in any relation to a representation – one only needs the relevant auditory (visual, recollection, sensory) capacities. It is perhaps slightly more controversial to say that one does not need a *private* language, or to stand in any kind of relation to a mental representation, in order to hear (see, remember, feel) that p. Although I would be willing to make this further claim, my discussion will not rely on this assumption.

We do not always have neat ways of expressing evidential attitudes in natural language. We generally have lots of propositional evidence about whether we are cold or thirsty, where our limbs are, whether we are moving, which way is up, and so on and so forth, which we have by our capacity for proprioception. In English, at least, we do not have any simple verbs for describing these states.

It should be clear that most evidence obtained in the ways detailed is inexact. In the clearest cases the evidence is inexact because the subjects eyesight is poor, or she only caught a short glimpse of something, or it was in the periphery of her visual field, or her memories had faded, and so on. But even when our sensory organs are all in good condition our evidence will be inexact. The amount of computing power we would need to calculate the exact path a tennis ball would follow on the information that it has been hit a certain way is typically too high for most humans (even more so when the examples become more complicated, such as spilling a bag of rice.) However we are generally able to do rough and ready estimations, without any conscious calculation, which give us a rough idea of where the ball will land. My best evidence is not that the ball will land exactly at location l or that it will land within the circular region r of diameter 3.8m or whatever, it will rather be inexact in the way characteristic of the examples we have been discussing. Most of our conscious and unconscious decisions are based on inexact information of this nature and understanding this information is crucial if we are to apply these epistemological considerations beyond the contrived decision problems, typical in formal epistemology, to commonplace decisions such as whether to hit a tennis ball when it looks like it might be going out.

Let's now focus on a particular example. Suppose that, after seeing Harry for the first time, I gain some evidence about how much hair he has. I can see that he has *some* hair, but as usual my evidence is inexact. I claim that

$$\text{For for some, but not all } n\text{, can I see that Harry has less than } n \text{ hairs} \tag{2.1}$$

As with all the the evidential attitudes we have discussed, 'sees that' is a propositional attitude: syntactically p in 'S sees that p' can be substituted for any grammatical sentence, and semantically we may view 'sees' as expressing a relation between a subject and a proposition. We may thus ask: what is the *strongest* proposition I stand in the seeing

relation to in this situation?[5] One may similarly ask what the *strongest* proposition that my evidence fully supports in this situation is. Assuming that the only evidence I have in this case is my perceptual evidence we should expect the answers to both these questions to be the same.

I shall consider two answers to this question

1. For some n, the strongest proposition about Harry's head that my evidence fully supports is the proposition that Harry has at most n hairs.

2. The strongest proposition about Harry's head that my evidence fully supports is a vague proposition about the number of hairs Harry has.

Before we tackle this question a few clarifications are in order. By the strongest proposition about Harry's head that my evidence supports I mean a proposition about Harry's head which my evidence supports and which entails all other propositions about Harry's head that my evidence supports. Assuming that one's evidence is closed under logical consequence, one can show that there always is a strongest proposition that is part of one's evidence: the conjunction of all the propositions supported by your evidence. Similarly, (assuming that the propositions about Harry's head are closed under conjunctions) there will always be a stongest proposition about Harry's head that my evidence supports.

The notion of entailment between propositions cannot be taken to be strict implication if we are to include vague propositions in this algebra. I shall assume, as per usual, that whether someone is bald or not supervenes on the precise facts about hair number, distribution and colour. To spare myself a few words, I shall simplify things further by assuming that it whether someone is bald only depends on hair number. Under our simplification, it follows that for some n it is necessary that Harry is bald if and only if Harry has less than n hairs – although it is vague for which n this holds. This notion of entailment is useless for epistemic matters such as evaluating your evidence, for it is exactly these necessities we cannot know because of vagueness. We shall need to avail ourselves of the more epistemic notion of entailment outlined in chapter 1, i.e. p entails q just in case every (im)possible p world is a q world.

It seems clear that the propositions stating that Harry has n hairs, where n ranges over numbers, are linearly ordered by entailment: that Harry has less than n hairs entails that he has less than m hairs whenever $m \geq n$. Where does the proposition that Harry is bald fall in this ordering? The answer, of course, is that doesn't. It appears below (i.e. it entails) some of the precise propositions, including the proposition that Harry has 10^{10} hairs, and appears above (is entailed by) others, including the proposition that Harry has 0 hairs. But if it's vague whether someone with N hairs is bald, and we know that Harry has exactly N hairs we cannot know whether Harry is bald. It's thus epistemically possible that Harry is not bald and has N hairs, so the proposition that Harry has N hairs does not entail the proposition that Harry is bald. Conversely, if I know Harry is bald, I do not know he has N hairs for that would be to know the vague, so it would be epistemically possible that Harry is bald and does not have N hairs, establishing that the proposition that Harry is bald does not entail that Harry has N hairs.

[5] In general it is odd to say that we see, if we are rational, every logical consequence of what we see. But at least in this case it is natural to think that I've seen most of the relevant propositions about Harry's hair number that are logical consequences of the propositions I've seen.

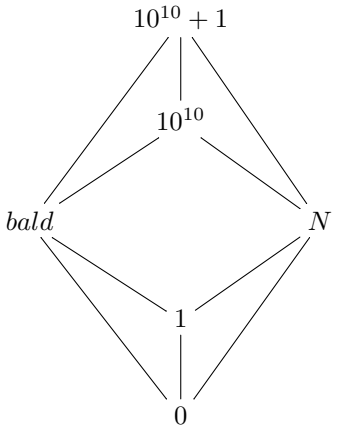
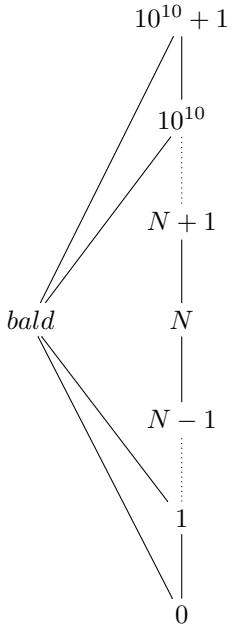


Figure 2.1: The proposition that Harry is bald neither entails nor is entailed by the proposition that Harry has N hairs



Despite the fact that certain vague propositions are independent of the precise propositions entailment-wise, they are not probabilistically independent of one another. Suppose that the probabilities representing my credences are initially uniformly distributed over the possible numbers of hair Harry might have. What are the probabilities of these heights given that Harry is bald? The precise propositions entailed by the proposition that Harry is bald now have probability 1 given that he's bald. However, the probabilities conditional on Harry's baldness will presumably drop smoothly: for each n, if the probability that Harry has exactly n hairs is x then the probability that Harry has exactly $n+1$ hairs will be just below, or the same as x. Thus if N is the largest number such that the proposition that Harry is bald entails that Harry has at most N hairs then, although the proposition that Harry has at most $N+1$ hairs neither entails nor is entailed by the proposition that Harry is bald, it is probabilistically supported by it, in the sense that its relatively low

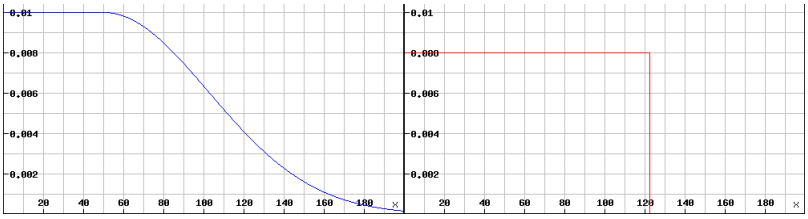


Figure 2.2: A smooth curve and a sharp curve of n against credence that Harry has n hairs.

initial probability will increase so that it is almost 1 on the supposition that Harry is bald. If I learn that Harry is bald the probability in some of the propositions not entailed by the proposition that Harry is bald will increase.

Formally, p is probabilistically independent of q for S if $Cr(p \mid q) = Cr(p)$ where Cr represents S's credences – the foregoing demonstrates that the vague and precise propositions are not probabilistically independent for a rational agent with my evidence. In my view this kind of probabilistic dependence is no accident; if your credences about Harry's hair number and whether he is bald are probabilistically independent in this situation you are being conceptually incoherent. This can be contrasted with a certain kind of Bayesian permissivism in which all probability functions represent a possible coherent assignment of degrees of belief. For this Bayesian there will be rational priors allow these propositions to be probabilistically independent.

If one were to begin with evenly distributed credences and were to update instead on the proposition that Harry has less than n hairs, one would have a sharp probability function: $Cr(Bk) = \frac{1}{n}$ for $k \leq n$ and $Cr(Bk) = 0$ otherwise. The smooth curve is intuitively what you'd expect to be the right one, and, I would conjecture, is closer to the curve that corresponds to the credences people usually adopt after having learnt from inexact experience. Sometimes we can tell whether or not someone has some hair from a cursory glance. In such a case the above graphs would need to be modified. Let us assume, however, that Harry is very bald, so that I can't rule out that he has no hairs at all and that my prior credences are uniform.[6] In such a case, one would expect one's posterior credences, after undergoing the inexact experience, to be closer to the blue graph in figure 2.2, as opposed to the red one. There is thus a striking resemblance between the effect of updating on a vague proposition and the effect of undergoing an inexact experience.

One way to see whether someone's credences correspond to the blue or red curve is to look at that persons betting behaviour. If, for each $k \leq n$, she would be willing to accept a bet that cost £1 and payed out £n if Harry had at most k hairs, and she would not accept any bets of this form if $k > n$, then the agent's credences are best described as conforming to the red graph in figure 2.2. The agents betting dispositions change suddenly; she is willing to pay anything up to a pound for a bet that Harry has at most n hairs, but is unwilling to pay anything for a bet that Harry has at most $n + 1$ hairs. On the other hand, if her credence corresponded to the blue curve in figure 2.2, then the amount she would be willing to pay for a bet that payed out a constant amount if Harry had at most k hairs would decrease continuously as k increases. If my visual experience of Harry's head is inexact then it should clear that I would become less and less certain that Harry had at most k hairs as k increases, which is observably born out in the kinds of bets I would accept.

In these kinds of situations described it should be clear that betting behaviour conforms to the latter kind of credence distribution. Whether, in ordinary situations, the agents credence is identical to the the credence your evidence recommends (i.e. the evidential

[6]Have uniform priors is also an unrealistic assumption – I presumably assign lower credences to extremely high numbers. It would be possible, but more complicated, to run the example without the assumption of uniform priors.

probability) is another matter altogether. However, the fact that we do appear to have smooth credences instead of sharp credences, and we are not criticisable for doing so, is quite suggestive.

2.2 Updating on your evidence

How is the evidential probability that Harry has n hairs determined from my evidence when I am in such and such a state? In this section I shall argue that extant approaches to updating do not adequately answer this question. I shall argue that Jeffrey conditioning relative to some non-propositional evidence, such as a visual experience, does not provide an answer to this question, and that conditioning on a precise piece of propositional evidence does not provide a *plausible* answer to this question.

Let us begin with the first approach. A purportedly distinct way of revising your credences in response to inexact evidence was proposed by Richard Jeffrey in [63]. We shall see, however, that Jeffrey's generalisation of conditionalisation does not tell us how we ought to respond to inexact evidence. We shall see, in fact, that our own theory of inexact evidence is not an alternative to Jeffrey's theory but a supplementation of it.

In the simplest case in which one's credence in a particular proposition e is raised by some experience, Jeffrey's theory states that your posterior credence, Cr', and your prior credence Cr must be related in the following way:

$$Cr'(p) = Cr'(e)Cr(p \mid e) + Cr'(\neg e)Cr(p \mid \neg e) \tag{2.2}$$

We can consider e to be some piece of 'evidence' that you have become less than certain in. Jeffrey conditionalisation (hence forth, JC) thus generalises the relation that holds between your prior and posterior credences that obtains when one updates by conditionalising on e (one can see this by setting $Cr'(e)$ to 1.) This generalisation allows one to account for changes in credences arising from what Jeffrey calls 'uncertain evidence': changes in an agents evidential state which, according to Jeffrey, does not involve the agent becoming certain in any proposition. Such cases might perhaps include learning from inexact evidence.

In the simplest case, e may be some proposition whose rational credence is determined by your new experience and background beliefs. However, it might be that your credences over a larger partition, $\{e_1, \ldots, e_n\}$, are determined in this way. In which case we may generalise 2.2 to:

$$Cr'(p) = \sum_{i \leq n} Cr'(e_i)Cr(p \mid e_i) \tag{2.3}$$

If e_i is the proposition that Harry has i hairs then it is trivial to find an instance of Jeffrey's equation 2.3 which matches the smooth curve in figure 2.2, provided none of the e_i have a prior credence of 0 (one simply sets the coefficients $Cr'(e_i)$ to the values of the graph.) We must be clear from the start which problem Jeffrey's extension of conditionalisation is supposed to solve. In Jeffrey's own words

> The problem is this. Given that a passage of experience has led the agent to change his degrees of beliefs in certain propositions $B_1, B_2, \ldots, B_n$ from their original values, $Cr(e_1), Cr(e_2), \ldots, Cr(e_n)$ to new values, $Cr'(e_1), Cr'(e_2), \ldots, Cr'(e_n)$ how should these changes be propagated over the rest of the structure of his beliefs? If the original probability measure was Cr, and the new one is Cr', and if p is a proposition in the agents preference ranking but is not one of the n propositions whose probabilities were directly affected by the passage of experience, how should $Cr'(p)$ be determined? [Notation has been modified to fit current conventions.] [63]

It should be noted that formally any proposition can be substituted for e in the first equation, and any partition in the second. For Jeffrey propositions $e_1, \ldots, e_n$ are those propositions 'whose probabilities [are] directly affected by the passage of experience.' What might this mean? A natural thought is that the propositions, $e_1 \ldots e_n$, are those propositions whose credences are rationally determined by your background credences and the fact that you've had a certain experience. However, if any partition of propositions posterior credence is rationally determined by your new experience and your prior credences, all partitions of propositions are. This is, after all, what JC itself says: once your posteriors over a given partition are fixed, the rest of your credence ought to be determined from this by Jeffrey conditionalisation.

If we assume that evidence is propositional, and that we update by conditioning on our evidence then it is possible to state what contribution one's evidence makes to one's doxastic state in a way that does not depend on one's prior beliefs. Within Jeffrey's framework, in which it is often assumed that evidence is non-propositional, it is not so easy to separate out the contribution a piece of evidence makes from the contribution your prior credences make: the values of $Cr'(e_i)$ are determined by your prior credences as well as the experience you have undergone. There is clearly no universally applicable answer to the question 'what must your posterior credences in e be after such and such an experience type.' For some people e might be initially highly probable whereas for others it may be very improbable, so it should be clear that no type of experience can be guaranteed to bring any two people into agreement about e. What one would want instead is some correlation between experiences and propositions telling you how much those experiences support those propositions, from which one can determine, given someone's background credences, how much posterior credence *they* should assign to that proposition – JC alone does not do this. This is called the 'input problem', and as yet there is no particularly satisfactory answer to it (see the exchange [35] [49] for a failed solution to the input problem.) Without a solution to the input problem, JC has little content. It just states a relation that must hold between your prior and posterior credences. Furthermore, that relation is very liberal: under simplifying assumptions, if we choose a fine enough partition it can be shown that any transition between two probabilities that preserves certainty is permitted by the constraints of JC. This is perhaps too liberal; one might think that there are some changes of credential state consistent with the JC constraint that would not be permitted by any possible evidence you could obtain.[7] For example: suppose that prior to receiving visual experience v, resulting from observing a green cloth by candlelight, the proposition that the cloth is green and the proposition that number of stars is odd are probabilistically independent. Suppose also that after experiencing v, I become certain that the number of stars is odd. It seems that this kind of transition of credences simply isn't permitted by the kind of evidence that I've recieved.[8] I shall refer to a transition between two credences that *is* permitted by some possible evidence a 'realizable' Jeffrey conditioning. The case just discussed is an unrealizable Jeffrey transition.

Unlike conditionalisation, JC should not be read as telling you how to respond to your evidence, but should rather be read as telling you how, once you have changed your credence in e to some degree, the rest of your credences redistribute to accomodate this change.[9] As a theory of inexact evidence, JC remains silent. I may increase my credence in p when I

[7] One reason to think this is that that the *order* in which you receive the evidence between two times shouldn't matter to your credence at the latter time. But if any Jeffrey conditioning counts as a rational transition then there are pairs of Jeffrey conditions which result in different posterior depending on the order in which you update on them.

[8] It should be noted that as far as Jeffrey himself was concerned, this is a perfectly rational transition: how your credences respond to your experiences is a purely causal involuntary matter, about which rationality has nothing to say.

[9] Conditionalisation can be read as telling you how to redistribute your credences once you have become certain in e, but it is most naturally read as telling you *to become* certain in e (and redistribute accordingly) when e is part of your evidence. There is no simple connection between your evidence and Jeffrey conditioning in the same way.

have no new evidence for p, or even if all the evidence is against p, and still comply with JC by redistributing my credences correctly. Similarly, it is in full compliance with JC to make no change at all to my credences in p even when there is strong evidence for p. We want to know: when *should* I change my credences and how. When I see a tree in the distance, for all JC says, I may just keep my credences about its height the same even though I have new information.

Jeffrey's theory is thus not really a theory about how you ought to revise your credences in response to inexact evidence after all, nor is it a theory of what inexact evidence is. In order to have a satisfactory theory of inexact evidence, JC must be supplemented. In §2.1 I defended the claim that when your evidence is inexact, your evidence consists of a vague proposition, and that your credences ought to be the result of conditioning your priors on this proposition. This, I claim, is just the sort of supplementation JC requires. The following is thus another feature of the evidential role vague propositions have:

> Conditioning on a vague proposition is equivalent to a realizable Jeffrey conditioning on a partition of precisifications.

The partition of 'precisifications' is, in the most general case, the partition of maximally specific precise propositions. In practical contexts however, we may use the term 'precisifications' to denote a coarser partition; for example, the precisifications of the proposition that Harry is bald can be treated, for all in intents and purposes, as the set containing, for each n, the proposition that Harry has exactly n hairs. So our principle states that conditioning on the proposition that Harry is bald is equivalent to Jeffrey conditioning over the partition of propositions containing, for each n, the proposition that Harry has exactly n hairs.[10]

In section 2.1 we argued that the smooth blue graph in figure 2.2 could be obtained by conditioning on a vague proposition. The ideas in this section can now be applied to demonstrate that conditioning on a vague proposition would have this effect. To set up an example, suppose that I have looked out of the window and have seen a tree at a distance. According to the current proposal the strongest proposition that I have seen, and that is therefore part of my evidence, is a vague proposition. For the sake of argument, let us suppose it is the proposition that the tree is about 200cm. In order to represent this situation we shall model propositions as sets of epistemic possibilities (or (im)possible worlds.) As in chapter 1, these can thought of as certain kinds of maximally consistent propositions; all that really matters is that since we do not know whether a tree that is 220cm is about 200cm (because, let's suppose, it is vague) we require there to be an epistemic state in which the tree is 220cm and is about 200cm, and a state in which the tree is 220cm and it's not about 200cm. For every possible height of the tree, h, and for every epistemically possible cut off point for 'being about 200cm', $\pm c$cm, there will be an epistemic state where the trees' height is h, and the cut off point for being about 200cm is being within $c$cm of 200cm. The relevant epistemic states can thus be represented by conjunctions of propositions taken from the following two sets: {the proposition that the tree is $h$cm $\mid h \in \mathbb{R}$} and {the proposition that, necessarily, a tree is about 200cm iff it's between 200-$c$cm and 200+$c$cm$\mid c \in [0, 200]$}. There are therefore many more states than there would be if we had only countenanced precise possible worlds. Suppose, for mathematical simplicity, that we restrict the possible values c and h may take to some large but finite set, and that my credences are uniform over this set. After updating on the proposition that the tree is about 200cm the possible heights of the tree graphed against my posterior credences over the possible precise heights of the tree would form a triangular shape with a point at 200cm. In reality, however, my credences would not be uniform: I take it that the cut off point $c = 1cm$ is very unlikely, as is the cut off point $c = 200$. In reality we should therefore expect my credences to conform roughly to figure 2.3

[10]It should be noted that the converse to our claim – that every realizable Jeffrey conditioning over the space of precise propositions is the result of conditioning on a vague proposition – is not being considered

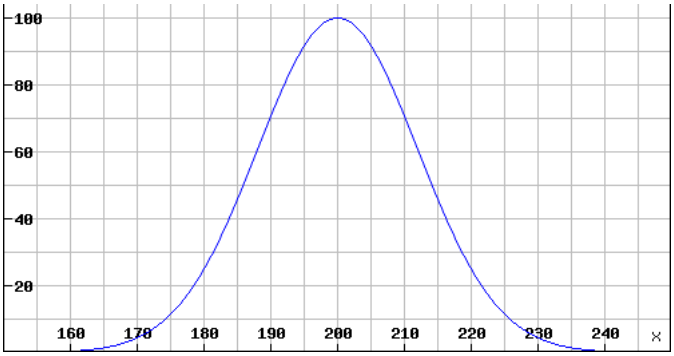


Figure 2.3: A graph of n against the proportion of epistemic states where the tree's height is $n$cm where it's also about 200cm.

Could the graph in figure 2.3 be obtained by conditioning if we had not assumed the theory of vague propositions defended in chapter 1? If the set of states were simply the set of propositions saying that the tree is exactly $x$cm tall, then all the propositions would be precise facts about which sets of numbers the trees height falls in. Presumably the strongest proposition in this algebra that you learn will be of the form: the tree is between $x$cm and $y$cm tall. If your priors about the trees height are uniform, the result of conditioning on a proposition like this will be rectangle instead of a smooth graph. In order to get a non-rectangular graph, as in figure 2.3, by conditioning on a proposition we need to add extra states to the model. But this proposal on its own seems mysterious; the crucial aspect of this proposal is that the extra states arise when we can draw distinctions within a single possible world about what vague matters are like there.

This fact marks a significant benefit of the theory of vague propositions I have been espousing. On a possible worlds theory, or any theory which postulates only precise propositions, there will simply be no proposition which can take you, as a result of conditioning on it, from having evenly distributed credences to the smooth credences represented in figure 2.3.

Neither can a possible worlds theory account for the following scenario. Suppose that Alice and Bob initially have identical evidential probabilites and that they both are looking at a tree in the distance, which, let us suppose, is 500cm high. As expected, their evidence is characteristically inexact. Suppose furthermore that in this situation they both have exactly the same precise evidence; the strongest precise proposition that is part of Alice and Bob's evidence is the proposition that the tree is between 400cm and 600cm (say.) However there is a difference: Alice's eyesight is better than Bob's. While neither Alice nor Bob can rule out that the tree is 401cm or 599cm, Alice's probability distribution forms a high peak at 500cm which quickly subsides remaining quite low at the edges (400cm and 600cm.) Bob's probabilities are much more evenly distributed, they increase steadily from 400cm, peaking at 500cm, and decreasing steadily until 600cm. This difference is easily explained by a difference in their propositional evidence on my view, but it cannot be explained by a difference in their precise evidence.

What this example shows is that one's evidential probability distribution does not supervene on one's precise evidence. Both Alice and Bob have the same precise evidence, yet they have different evidential probability distributions. In particular, this shows that evidential probability is not determined by conditioning on your precise evidence; thus either the agents evidential probability is not determined by conditioning on her evidence, or her evidence does not consist only of precise propositions. Since, as I have argued in this section, the view that we update on non-propositional evidence via Jeffrey conditioning

here.

does not solve our problem it seems that the best explanation relies on the existence vague propositional evidence.

2.3 The supervenience of vague beliefs on precise beliefs

Given we have some rough idea about how we ought to distribute our credences when our evidence is inexact the thesis that our inexact evidence is vague evidence puts some constraints on the evidential role of vague propositions. For example, it is presumably true that if my strongest evidence is the proposition that Harry is bald, then for some n, I'm rationally mandated, relative to my background beliefs, to be uncertain whether Harry has at most n hairs. But even if we keep fixed my prior credences about all the relevant precise facts, this thesis only entails that I should have a non-extremal credence. It is left open whether there is exactly one non-extremal credence I'm required to have, or whether there are several credences any one of which is permissible for me to have.

Conversely, if I'm rationally certain (i.e. my evidence entails) that Harry has exactly N hairs, where it's vague whether someone with N hairs is bald, then our discussion leaves it open what my credence in the proposition that Harry is bald should be. Even if we agree that I should be uncertain whether Harry is bald (because it is vague) there is still a range of credences strictly between 0 and 1 which I could have. Is there exactly one credence that I'm required to have, given that my evidence about the relevant precise facts is complete, or are there multiple credences it is permissible for me to adopt?

I favour the claim that you are required to have a particular credence. In this section I shall defend this by way of a more general supervenience thesis: that one's credences in the precise propositions, provided they are the rational ones to adopt on your evidence, determine what credences one ought to have in the vague propositions:

> Necessarily, for any pair of coherent priors, $Cr, Cr' \in Coh$, $Cr(p \mid w) = Cr'(p \mid w)$ for every maximally specific precise proposition w, and proposition p.

Many philosophers have asserted that vague propositions are somehow 'non-factual'. It has been pointed out that it is quite hard to pinpoint exactly what this amounts to if one accepts, as I have been, classical logic and the thesis that vagueness involves uncertainty. This is a topic we will come back to in more detail in the next chapter. However one feature of the picture I am endorsing, which is partially explicated by the above principle, is that vague beliefs are epiphenomena of one's precise beliefs. Our rational beliefs about the vague are completely determined by our rational beliefs about the precise. There is no flexibility about how to distribute your credences once you've learnt all the precise facts: rationality dictates how you then distribute your credences in the vague.

It is perhaps worth comparing vague propositions to other propositions that are often alleged to be, in some sense, non-substantive, or at least to have a secondary or derivative epistemological status. Certain propositions – propositions expressed by paradigm examples of 'analytic' sentences – are such that your credences in them supervene on your credences in the more substantial propositions; they are dictated, given your other credences, by very general norms of conceptual coherence. An example would be the proposition that all crimson things are red; here the supervenience is trivial since all conceptually coherent credences assign probability 1 in this proposition. Another example would be the proposition that there are eight solar planets and all vixens are female foxes. In this case, your credence is derivative on your credence in the proposition that there are eight solar planets, indeed, it ought to be identical to it. A much more controversial example, but one which is closer to our present concerns, involves conditionals: if the Ramsey identities are true, your credence in a conditional rationally supervenes on your credences in the antecedent and the conjunction of the antecedent and consequent.

The derivative status of our beliefs in these propositions is, in both cases, due to *conceptual* entailments between them and the substantive claims.[11] This marks a salient difference between these examples, and the type of supervenience beliefs in vague propositions have. We argued earlier that there are no conceptual entailments between the proposition that Harry is bald, and the proposition that Harry has N hairs. The relation between one's epistemic attitudes in the vague and the precise does not consist purely in entailments, but it is still a relation which determines your attitudes towards the vague uniquely, given your attitudes in the precise. Once one is omniscient about all the substantive non epistemologically derivative claims, our credences in the other propositions are fixed by norms of conceptual coherence. In the examples just considered this is just a matter of us having a credence of 1 or 0 in those propositions. For vague propositions, there it may be that some intermediate credence is the only conceptually coherent credence to have.

This explains one puzzling feature of vagueness, which is not shared by the propositions expressed by analytic sentences. Even if your evidence is as complete as it can be, you may still be rationally uncertain about the vague. I might, for example, know everything there is to know about Harry's hair number, distribution and colour, but not know whether he's bald. It is natural to think that if one is rationally uncertain about something, there must be some further evidence that could be obtained that would settle the matter. Explaining why my uncertainty about Harry's baldness is uneliminable is achievable on the current view: my uncertainty, given my particular evidence, is a *requirement of conceptual coherence.* It is part of the attitudinal role of that proposition, as it were, that I have a particular credence, x say, when I'm rationally certain on my evidence that Harry has this particular hair number and distribution.

Before we move on a few words of clarification are in order. The first point has to do with the term 'conceptual truth.' It is often applied to sentences or thoughts, and it is sometimes said that conceptual truths are those sentences or thoughts whose truth is guaranteed purely in virtue of the meanings of the words or concepts that the sentence or thought is composed of. The term 'conceptual truth' (and also the older term 'analytic truth') has become closely associated with the phenomenon I am concerned with: that for some propositions p (e.g. the proposition that all crimson things are red) it necessary that anyone who properly considers p would be automatically justified if she believed that p. I take it that this phenomenon may or may not be explained in the traditional way, in terms of our competence with words or concepts used to express these propositions. I shall, however, continue to use the word 'conceptual truth', as applied to propositions, (and 'conceptual entailment') to describe the phenomena.

Secondly I am using the word 'ought' objectively when I say that one ought to have a certain credence in p. For example, if p is a tautology that is widely disputed among philosophers, it might be reasonable for me as a philosopher to have some intermediate credence in p.[12] However if classical logic is correct, I do not in fact have any evidence against p and, indeed, p is necessarily and vacuously part of my evidence (or so I claim) so I am justified in believing it. I might even reasonably believe I have evidence *against* p, but since all evidence must at least be true, I'd be wrong about this. What I ought to believe is therefore objective and external to me and something I might reasonably fail to realise.

I shall henceforth use the term 'conceptual requirement' to denote both the simple kinds of requirements, such as having credence 1 in the proposition that vixens are female foxes, and the more complicated ones, such as having a certain credence in the proposition that Harry is bald, given you know he has N hairs. Again, I shall understand 'requirement'

[11]The term 'conceptual' is close enough to what I mean here, at least for the purposes of this chapter. However, whatever the special status of these entailments is, I'm inclined to think it does not extend to many of the alleged examples of *a priori* entailments; such as, say, mathematical entailments.

[12]This point is sometimes made vivid by noting there are many reasonable deviant logicians, but I think the point generalises even to classical logicians. While this thesis is a defense of classical logic and I therefore count myself as a 'classical logician' (at least when no empty names or modal operators are present) I do not have credence 1 in excluded middle and believe this is quite reasonable.

objectively – one may reasonably fail to know what you are required to believe. I shall use the term 'evidential role' for a complete profile of a propositions conceptual requirements.

2.3.1 The supervenience thesis

Let us now move on to defending the supervenience thesis.

We began this discussion by drawing an analogy between the kinds of requirements characteristic of propositions expressed by conceptual truths and the requirements characteristic of vague propositions. In fact it can be seen that these requirements are of one and the same type. Consider

All crimson things are red (2.4)

All auburn things are red (2.5)

(2.4) denotes a paradigm example of a conceptual truth. Therefore it is a conceptual requirement that I have full credence in the proposition that all crimson things are red. However (2.5) is not a conceptual truth: there are auburn things such that it's vague whether they're red. I therefore am not required to assign full credence to the claim that all auburn things are red.

The supervenience thesis entails that for each shade F there is some specific credence, x, that I'm supposed to have in the claim that all F things are red provided I know all the precise facts (let's say, all the facts about which objects are which precise shades.) When F is 'crimson' that number is 1. As F varies over the shades in between crimson and auburn the number cannot remain at 1, since we know that we should not be certain that all auburn things are red. According to the picture I prefer, conceptual requirements such as the requirement to have credence 1 in the proposition that crimson things are red transform seamlessly to the kinds of requirements I postulate – for example the requirement to have a credence of 0.745 in the proposition that auburn things are red. Is it possible that when F is substituted for 'auburn' there are multiple credences it is permissible for me to have?

On the simplest picture, the one I am trying to motivate, the answer is 'no.' Just as there is one specific credence I'm required to have in the proposition that all crimson things are red, there is one specific credence I'm required to have in the proposition that all auburn things are red.

The simplest picture also provides my answer to input problem for Jeffrey conditioning we mentioned in section 2.2: given a prior credence Pr, and evidence e (whether propositional or not), what partition and what coefficients should I Jeffrey condition relative to? In my set up e is a vague proposition and your posterior credence should be the result of conditioning on e. We argued that this was equivalent to Jeffrey conditioning on a partition of precise propositions –the partition of precisifications of e. The supervenience thesis provides us with an answer to what the coefficients are: the coefficient for the precisification p is the probability that every conceptually coherent prior assigns to e conditional on that precisification. This number is independent of the agents priors, as required.

What would it take for the simplest picture to be false? Could there be two rational agents who were certain about all the relevant precise matters, and yet disagreed about a vague proposition? It is very hard to see how this disagreement would manifest itself.

If two agents could disagree about a vague proposition p whilst having all the precise information available and having the same preferences how would this difference manifest itself. Could it manifest itself in a difference in behaviour? Could it manifest itself in some other aspect of the agents internal mental state?

Under certain assumptions a difference in one's credences in the vague really could manifest itself in a physical behavioural consequence. The example I have in mind runs as follows.

> Alice and Bob are in an auction house and they are both bidding on a relatively inexpensive vase which is a shade of turquoise that is indeterminate between being blue and green.
>
> Now suppose we know the following things about Alice and Bob. Firstly, they have exactly the same telic desires and have the same credences in the precise propositions. We also know the following facts. (1) as it happens they are both very peculiar people and they care intrinsically about owning green things. (2) although they both know exactly what precise shade the vase is, Alice is more confident than Bob that the vase is green.

In this scenario I think it is clear that Alice should be willing to bid more than Bob for the vase. Since they care about the same things its value is the same for both Bob and Alice. However, since they know that it's vague whether the vase is green they are uncertain whether it's green and therefore the purchase is a gamble on its being green. However, according to Alices subjective credences the odds that it is green are better than they would be according to Bob's. Therefore Alice ought to outbid Bob.

However there is a feature of this example which is particularly bizarre. In order for it to work both Alice and Bob have to care intrinsically about owning green things. Indeed, any example of a difference in behaviour due purely to differences in beliefs about the vague sees like it will have to involve agents who care intrinsically about the vague. There is something very odd about this, the full discussion of which I must delay to the next chapter.

However, if one thought that it was not possible to rationally care intrinsically about the vague then there is a serious problem for the supervenience deniers. If whenever two people are certain in all the precise facts they are behaviourally equivalent, then what role are the differing credences in vague propositions playing in the agents psychology? Notice that if they really are behaviourally equivalent even their linguistic behaviour will be the same: they will both report the same degree of confidence in p. If no failure of the supervenience thesis could manifest itself in any kind of behavioural difference it seems hard to imagine what theoretical importance differences like this might have.

2.4 A principle of plenitude for vague propositions

According to linguistic theories of vagueness vague propositions have a derivative status. Either there are no vague propositions or there are and they are parasitic on the vocabulary of the language. In either case there are at most countably many vague propositions entailing any particular maximally consistent precise proposition corresponding to the distinctions one can make in the language.[13] This dependence on the language seems to be particularly odd, and in chapter 1 we suggested that there were vague propositions not expressed by any sentence.

In chapter 1 I argued for the thesis that we shouldn't individuate propositions modally, or identify them with sets of possible worlds, we should rather individuate them by their attitudinal roles and we should identify them with sets of (im)possible worlds (maximally specific epistemic/bouletic/doxastic/etc states.) How should we express the idea that propositions exist independently of language and thought?

I propose the following principle of plenitude for propositions:

> For every possible evidential role there is some proposition with that evidential role.

Recall that the evidential role of a proposition is simply the set of its conceptual requirements. We may state the principle of plenitude rigorously as follows. Let P denote the

[13] In principle there could be $2^{\aleph_0}$ maximally *logically* consistent sets of sentences above a given set of precise sentences, however it is not clear that they all correspond to conceptually consistent propositions. (For example, there are $2^{\aleph_0}$ maximally consistent sets extending Peano arithmetic, but most of these are not genuinely conceptually consistent because they are ω-inconsistent.)

set of maximally specific precise propositions (i.e. those precise propositions entailing all precise propositions which entail them.) Then an evidential role will be represented by a function $E : P \to [0, 1]$, telling you what your credence should be if you were to learn a given maximal precise proposition

> The Principle of Plenitude
>
> Let E be an evidential role. Then there is some proposition, p, such that for every coherent prior Cr and precise proposition $w \in P$, $Cr(p \mid w) = E(w)$.

As an example, the principle of plenitude entails that there is a proposition p such that whenever one is rationally certain that Harry has N hairs, one's credence in p ought to be 0.798 (set $E(w) = 0.798$ whenever w entails that Harry has N hairs.) The principle of plenitude entails the existence of an abundance of propositions. Note that while the evidential roles generate propositions, they do not individuate them. For example, by Plenitude there is a proposition, p, such that one is required to have credence $\frac{1}{2}$ in it conditional on any maximally specific precise proposition. It follows by probability theory that $\neg p$ has the same evidential role as p, yet these propositions are always distinct in a Boolean algebra.

A vague propositions evidential role is not just determined by its interaction with precise propositions. For example, the proposition that the ball is blue and the proposition that the ball is turquoise are evidentially relevant to one another even if we know all the precise facts. Suppose that we know which exact shade the ball is, and that that shade is such that it's vague whether it's turquoise and blue, just turquoise or just blue. Although having further evidence that it was turquiose would be impossible, we can still ask if the proposition that the ball is turquoise is evidentially relevant by considering the conditional probabilities (even if these probabilities could not represent the rational credences of an agent on her evidence.) And it is: the probability that it's blue presumably decreases on the hypothesis that the ball is turquoise since we have to rule out the blue and not turquoise possibilities, and there are more just turquoise than turquoise and blue possibilities. The principle of plenitude could therefore be extended to entail connections like these with the following impredicative version of the principle:

> For any proposition, p, and pair of functions $E_1, E_2 : P \to [0, 1]$ there is a proposition q such that for any $w \in P$, $Cr(q \mid p \wedge w) = E_1(w)$ if $Cr(p \mid w) > 0$ and $Cr(q \mid \neg p \wedge w) = E_2(w)$ if $Cr(\neg p \mid w) > 0$, where Cr is a conceptually coherent prior.

The principle of plenitude is a special case of this principle which can be seen by setting p to $\top$. It is impredicative because any proposition q that is generated by this principle can be fed back in as p. In the appendix it is shown that both principles are consistent.

It is tempting to ask why one ought to have a particular credence, 0.798 say, in the proposition that Harry is bald given one knows the relevant facts about his head. This is simply not a question with a reasonable answer on this way of individuating propositions. Part of what it is to be the proposition that Harry is bald is to have the specific epistemic profile it in fact has, which includes this conceptual requirement. It is like asking, on the modal way of individuating propositions, why the proposition that Harry is bald corresponds to one set of worlds and not another.

However, the thought which I take it this question is trying to latch on to might be better expressed by the question: why does the sentence 'Harry is bald' denote a proposition with one set of conceptual requirements rather than another? This is presumably a question of metasemantics and a question whose answer is vague. The appearance of precision is really an illusion. A similar problem arises for the modal way of individuating propositions. One might ask why the proposition that Harry is bald contains worlds where Harry has 784 hairs but not worlds where he has 785 hairs. On the modal way of individuating propositions this is just a bad question: it amounts to asking why a given set has the members it in fact has as members. The sensible question in the vicinity is why does the sentence 'Harry is

bald' express this proposition and not another, or why are we justified in calling this set of worlds 'the proposition that Harry is bald'. Both are metasemantic questions, and one should expect there to be vagueness concerning which propositions are picked out by which sentences.

Recall the example of the tree seen from a distance. Suppose that initially my credences are uniform over the possible heights less than 1000cm. After the evidence is in my credences form the curve displayed in figure 2.3. What proposition would have the effect that updating this way would cause this change in credence? Here the principle of plenitude kicks in. According to the principle there will be a proposition whose evidential profile ensures that learning it will change my uniform starting credences to the curve. Let that curve be represented by the function Pr. For simplicity lets say the partition of precisifications is finite, so lets pretend the maximally specific precise propositions are the propositions saying that the tree is $n$cm tall (this simplification is harmless.) Let E be the evidential role given by $E(p_n) = Pr(p_n)$. Then the principle of plenitude entails there is a proposition, p, with evidential role E. In particular, if Cr is a coherent prior and e is the agents total evidence, so that $Cr(\cdot \mid e)$ is my credence before the observing and is thus uniform over the partion p_n for $n \leq 1000$, then conditioning on p will result in a curve like that in figure 3: $Cr(p_n \mid e \wedge p) = E(p_n)$. For proof the appendix.

2.5 Further issues and clarifications

Evidence. The picture of evidence espoused here does not sit well with certain internalist conceptions of evidence. According to the view I've defended one cannot have evidence that Harry is bald (or evidence that Harry is not bald) if it's borderline whether Harry is bald. But, one might think, surely I could have misleading evidence that Harry has exactly no hairs when in fact he has a borderline number. For example, if I initially have no idea how many hairs Harry has and a normally reliable person tells me that Harry has no hairs when in fact he has a borderline number, some philosophers would say I have evidence that Harry is bald even though this is a borderline proposition. According to other philosophers, however, this is not so: evidence must be factive in the sense that your evidence consists only in true propositions. In the above situation, for example, your evidence would include the proposition that someone has told you that Harry has no hairs, but would not include the proposition that Harry has no hairs since it is false. Your belief might be justified by your evidence in the sense that it is a belief with a high evidential probability, but it would not be evidence in the strictest sense.

I suggest that according to this conception one's evidence consists only of *determinately* true propositions. Imagine two parallel cases. In one case the reliable person tells you that Harry is bald instead of telling you he has no hairs, in the other she tells you he's not bald. Either Harry is bald or he isn't, so in one of the cases you have been told a true proposition. Yet on the externalist conception of evidence it is natural to think that, even though in one case the proposition you're told is true, in neither case is the proposition you are told evidence.

Rational belief If correct, the view outlined makes it clear why we cannot rationally have certain beliefs – for example why, given certain values of N, we cannot rationally believe that Harry has N hairs and is bald. The proposition that Harry is bald exists only to play a certain evidential role: to be your evidence when your evidence is gained through certain inexact mechanisms. When your strongest evidence concerning Harry's head is that he's bald, then when N and $N-1$ are in the borderline region your credence that he has N hairs is less than your credence that he has $N-1$ hairs. But by a standard piece of Bayesian reasoning (namely, Bayes' theorem) this can only happen if your credence that he's bald given he has N hairs is smaller than your credence that he's bald given he has $N-1$ hairs; thus your credence that he's bald given he's got N hairs is strictly less than 1. A symmetric argument would establish that your conditional credence in this situation must be greater

than 0 – your credence that Harry is bald given he has N hairs is intermediate.

A common objection to classical views that claim that it's impossible to rationally believe that Harry is bald whilst also being borderline bald invokes the thought that, since he's either bald or he isn't, God could come down and tell you which, making your belief rational. However the way this objection is usually put is vastly underspecified. What exactly is the situation being envisaged? How do we know the person telling us this is God?

The naïve image the objection conjures up is that of a glowing bearded fellow walking into the room, claiming to be God, and telling you that, even though it's borderline, Harry is in fact bald. However I think that no-one would claim that this would be sufficient to rationally justify your having the belief that Harry is bald and borderline bald. Imagine that he had instead told you that it was raining and that it wasn't, or that there are male vixens. In the latter case the rational thing to do given your evidence (assuming that you're certain God wouldn't lie, that the man did in fact say this, and so on) is to conclude that this person isn't God after all. In the original case the rational thing to do, assuming you're originally pretty sure the person is God, would be set your credence that the guy is God to some lower intermediate value (for the reason why it's not 0 in this case see the footnote.[14]) (Note that there might be scenarios where it's *reasonable* to believe that there are true contradictions. For example, if Frege, Gödel, Tarski and all the greatest logicians walked in and told you this it might well be reasonable to believe that there are true contradictions, but it is not rational on your evidence, in the strictest sense, to believe this since, assuming there are no true contradictions, you do not have evidence that there are true contradictions.)

Knowledge. Finally it is often argued that it is not possible to know p when p is borderline. The two points above go some way to explaining this intuition.

For example, many people would agree with the principle about not rationally believing propositions whose borderliness is *a priori*, and yet allow that one could believe, maybe even rationally, a contingently borderline proposition – that Harry is bald, for example – by being told that Harry is bald by a reliable person. Indeed, as we noted before, the reliable informer might even be correct about whether Harry is bald. Yet, despite the fact that Harry is bald, and the fact that you believe it quite reasonably, the thought goes that you still do not know that he's bald because it's vague. How do we explain ignorance in this case from the facts described above?

The first thing to note is that in this case you do not in fact have the evidence that Harry is bald. So even though your belief is true and justified by (i.e. has high evidential probability on) your evidence it's not clear that it constitutes knowledge. That said, some reliablists (see Lasonen-Aarnio [78]) have argued that a stubborn belief that is reliable, even if not rational, can still constitute knowledge. In the case of knowledge, as opposed to rational belief, it seems that a variant of the God example above could take a different turn. For example suppose that God told you that Harry was bald, but instead of coming to lessen my confidence that this guy is God I irrationally decide to believe everything he says. Then, perhaps, my stubborn beliefs would constitute knowledge, since, as it happens, the man speaking to me really is God and is a reliable source of information. A less fanciful example might involve two people, one of whom stubbornly forms a belief that Harry is bald, the other who forms a belief that he is not bald. Assuming that there's no generally applicable method they used to form these beliefs, it's natural to think that one of them has a reliable belief (while the other has an anti-reliable belief.) If this is all it takes to have knowledge, perhaps one of them knows whether Harry is bald.

[14] Assume that you already know that Harry has N hairs, and that we are therefore required to have a credence of $\frac{1}{2}$ that Harry is bald. Let G be the proposition that the guy is God, B the proposition that Harry is bald and has exactly N hairs. By assumption we have $Cr(B) = \frac{1}{2}$ and since we are initially pretty sure the guy is God we have $\frac{1}{2} > Cr(B \wedge G) > 0$. Since God is always truthful and has said B, we can assume that $Cr(B \mid G) = 1$. So $1 = \frac{Cr(B \wedge G)}{Cr(G)}$, so $Cr(G) = Cr(B \wedge G) \in (0, \frac{1}{2}]$.

Closure under conditioning. A natural principle states that rational credences are closed under conditioning: if Cr represents credences of a rational agent, then the result of conditioning Cr on a proposition with non-zero probability is also the credence of a rational agent. After all, if your credence is Cr, and $Cr(p) > 0$, then surely if you ended up *finding out* p it would be rationally required to condition on p? Sure surely it must be permissible to have credence $Cr(\cdot \mid p)$

Neither the principle nor the explanation hold on the view I am proposing. The principle is false for the following reason. Anyone who is rationally certain that Harry has N hairs is required to have a specific credence, $\alpha \in (0, 1)$, in the proposition that Harry is bald. Since it is possible to have evidence that Harry has N hairs, there is a rational credence function assign 1 to the proposition that Harry has N hairs. In this case the credence that Harry is bald should be α and therefore greater than 0. If we were to condition this function on the proposition that Harry was bald we would have a function which assigned a credence of 1 both to the proposition that Harry has N hairs and to the proposition that Harry is bald. This function cannot be rational since all rational credences assigning credence 1 to the proposition that Harry has N hairs assigns $\alpha < 1$ to the proposition that Harry is bald.

Counterexamples to conditioning work in the other direction as well. It is possible that one's strongest evidence is the proposition that Harry is bald. In which case, there are hair numbers within the border region such that you are not certain Harry has them – so in these situations your are not completely certain that Harry is not borderline bald. But the result of conditioning on the proposition that he's borderline bald would make you certain he was bald and borderline bald.

These features are not specific to my view, however. It is independently plausible that anyone who is certain that it's vague whether Harry is bald should be uncertain whether Harry is bald. Yet if Cr is a rational credence function assigning 1 to the proposition that it's vague whether Harry is bald, then the result of conditioning it on the proposition that Harry is bald will be irrational since it would represent an agent being certain that it's vague whether Harry is bald without being uncertain whether he's bald.

So, on independent grounds, the explanation for the closure principle must be faulty. It is easy to see why the explanation fails according to the externalist conception of evidence we have be espousing. Anyone who has evidence that Harry has N hairs (or, more generally, has evidence that Harry is borderline bald) cannot then obtain evidence that Harry is bald, since one's evidence must consist of determinately true propositions. In the other direction: anyone who has evidence that Harry is bald cannot then obtain evidence that Harry has N hairs, since if you have evidence that Harry is bald then he must be determinately bald, and thus have strictly less than N hairs. Since evidence is factive, he cannot also have evidence that Harry has N hairs.

Agreeing about the precise. Recall the supervenience principle we stated in section 2.3:

> Necessarily, for any conceptually coherent priors, $Pr, Pr' \in P$, $Pr(p \mid w) = Pr'(p \mid w)$ for every maximally specific precise proposition w, and proposition p.

This states a condition on rational priors. However, given our conception evidence as always consisting of determinately true propositions, the above principle entails the following principle concerning non-initial rational credences:

> Necessarily, for any rational credences, Pr, Pr', and maximally specific precise proposition w, if $Pr(w) = Pr'(w) = 1$, then $Pr(p) = Pr'(p)$ for every proposition p.

In other words, any two people who know all the precise facts should have the same credences in every proposition. This is because if one's evidence is a maximally specific precise proposition, then it is one's *strongest* evidence on pain of not being determinately true. Thus $Cr(p) = Pr(\cdot \mid w)$ for some coherent prior Pr and maximally specific precise proposition w and $Cr' = Pr'(\cdot \mid w)$. Thus by the supervenience principle $Cr = Cr'$.

One might think that the supervenience principle also entails the following stronger principle:

(*) Necessarily, for any two rational agents with credences Cr, Cr', if $Cr(p) = Cr'(p)$ for each precise proposition, then $Cr = Cr'$.

In other words: any two people who agree about the precise agree about everything. Here's why. Suppose that Cr and Cr' are rational credences (rational priors conditional on possible evidence), and suppose that $Cr(p) = Cr'(p)$ for every precise proposition p. Let w range over the set of maximally specific precise propositions.[15] Then:

1. $Cr(q) = \sum_w Cr(q \mid w).Cr(w)$ by probability theory.
2. $= \sum_w Cr(q \mid w).Cr'(w)$ since w is precise, and there is agreement on the precise.
3. $= \sum_w Cr'(q \mid w).Cr'(w)$ since, by my supervenience principle, $Cr(q \mid w) = Cr'(q \mid w)$.
4. $= Cr'(q)$ by probability theory.

(*) say that any two people who agree about all the precise things agree simpliciter. However (*) seems to be false. Suppose that Martha and George have exactly the same credences in the precise propositions concerning Harry's head: let's suppose both their credences conform to the blue curve in figure 2. However they came to these credences in different ways: George caught a brief glimpse of Harry's head, and the strongest proposition he learnt was a vague proposition, Martha learnt that Harry's hair was going to be cut to some number of hairs via a chancy event which happens to represented by the blue curve in figure 2. Although Martha and George agree about the precise they *do not* agree about everything: George, for example, knows that Harry is bald and Martha doesn't.

The problem is that strengthened supervenience principle is false. The error in its derivation is in step 3. The supervenience principle does not entail $Cr(q \mid w) = Cr'(q \mid w)$ unless Cr and Cr' are rational *priors*. For e.g. in the Martha George problem, George's credence is not a rational prior (it judges that Harry is bald. He could not rationally learn that Harry had N hairs.)[16]

Degrees of truth. Modulo the issues we discussed in chapter 1, maximally precise propositions in this framework are approximations of possible worlds in an intensional framework. Assuming this analogy, vague propositions then determine a function from possible worlds to real values in $[0, 1]$: the credence any conceptually coherent agent ought to have if she learnt that she was in that possible world.

Another approach to vagueness also identifies with each vague proposition a function from possible worlds to real numbers in $[0, 1]$. According to this view, the number a proposition is assigned at such a world is its truth value there. These theorists often espouse a non-classical fuzzy logic, due to the way in which the truth values at a world interact with the connectives. However other theorists keep classical logic and are close in spirit to the current view. Edgington [27], for example, espouses a view in which the truth values behave like probability functions (see also Kamp [66], Lewis [80] and Williams [136], for versions and developments of the view in a supervaluationist setting.) One might wonder how fuzzy views, especially Edgington's view, differ from the view defended here. Edgington, for one, thinks that there is a difference between truth values (which she calls 'verities') and credences. In [27] she writes:

> "Some philosophers (e.g. Williamson 1994) hold that vagueness is a species of epistemic uncertainty: there is a precise line which divides the red from the non-red, etc., but it is epistemically inaccessible to us. Were this true, verity would

[15] I shall bracket issues arising if there are uncountably many maximally specific precise propositions.

[16] You might be able to tell a consistent story in which Martha's credences are rational priors I suppose.

> be a kind of credence: the credence that a person with no relevant ignorance other than about the precise line, would give to a statement like 'that's red'. If a is redder than b, but neither is clearly (certainly) red, then one must, if rational, be more confident that a is red than that b is: a is more likely to be above the mystery line than b is. The nearer to clearly red is the nearer to certainly red. Credence and verity, I argued, have the same logical structure. This could be interpreted as grist for the epistemicist's mill. What better explanation of the analogy I have developed, than that verity is credence, and so vagueness a kind of epistemic uncertainty?"

The view I am defending is not epistemicism, but it agrees with the epistemicist in respects important to this discussion. According to Edgington verities and credences play different roles. Credences inform a rational person's actions, whereas verities do not. In support of this she notes the distinction between the following kinds of justifications:

> I prefer A to B by a long way. Therefore I prefer A with certainty to A or B happening with equal uncertainty, i.e. probability $\frac{1}{2}$, and I should prefer A or B happening with equal uncertainty to B with certainty.

> I prefer A to B by a long way. Therefore I prefer A with verity 1 to A or B each with equal intermediate verity, $\frac{1}{2}$, and I should prefer A or B with equal verity to B with verity 1.

The former principle is a valid principle in most decision theories, yet Edgington gives an example, in which whether A or B is borderline figures in my preferences, to demonstrate that the second is not a universally correct principle of decision making.

One might have the converse worry: that the credence a coherent person should assign to a proposition conditional on being in a world plays exactly the same role as a truth value.[17] Consider the following 'truth norm': if p is true then one should believe that p (and one should not otherwise.) One can formulate similar truth norms for partial beliefs, in which credences closer to the truth are better (see [64].) Let $E(w)$ represent the credence you should have in the proposition that Harry is bald conditional on knowing that you're in w. Then $E(w)$, qua the thing that your credences ought to match when w obtains, seems to play the very same role the truth value of the proposition that Harry is bald plays as described above. Is $E(w)$, the evidential role of the proposition that Harry is bald, not just the truth value of the proposition that Harry is bald had w obtained?

Other than Edgington's argument above (which I find convincing) I think there are two further points to be made. The first has to do with the truth norm. I think that in the sense in which we use the word 'should' in epistemology it's simply not the case that Copernicus should have believed that space-time is non-Euclidean; all his evidence pointed against this so there's a clear sense in which he *shouldn't* have believed this, even though it's true. However, the word 'should' is clearly context sensitive, and with a bit of scene setting it may well be possible to find contexts in which the truth norm holds. However *it is far from clear* that in the contexts in which the truth norm holds my credence that Harry is bald conditional on w, should-in-this-context be $E(w)$, and not 1 or 0 (depending on whether Harry is bald at w.) After all, the view that's being espoused is a bivalent one, and if it's possible to read 'should' in this unintuitive ultra-objective way it is natural to think that *all* my credences should be 1 or 0. Thus, on this reading of 'should', the disquotational (i.e. bivalent) truth values play the truth rule.

The second point I want to make involves an analogy with chances. Note the similarity between the following two claims

> For each proposition, p there is a function, E, such that for each maximally strong precise proposition w, and rational prior credence function Cr, $Cr(p \mid w) = E(w)$.

[17] I am indebted to Robbie Williams for pressing me on this objection.

For each proposition, p there is a function, F, such that for each maximally strong piece of admissible evidence at t, w, and rational prior credence function Cr, $Cr(p \mid w) = F(w)$.

In the both cases there is a special partition of the space of propositions such that every rational prior agrees with every other prior conditional on any element of that partition. The former principle is a consequence of the supervenience principle, the latter a consequence of the principal principle (see [81].) In the latter case $F(w)$ simply represents the chance that p at time t. In whatever sense E plays the truth role for the proposition p (relative to the partition of maximally specific precise propositions) then note that so do chances (relative to the partition of worlds with the same admissible evidence at t.) Since we are not at all tempted to call chances verities on the basis of truth norms alone, I think that we should not be tempted to call credences verities.

2.6 Appendix

The principle of plenitude says that, for any function t from the set of maximally specific precise propositions to $[0, 1]$ there is a proposition p_t such that for every conceptually coherent prior Pr, and maximally specific precise proposition, w, $Pr(p_t \mid w) = t(w)$ provided $Pr(w) > 0$.

As with any existence postulate, such Lewis's plenitude principle for possible worlds or the naïve comprehension principle in set theory, there is a question about its consistency. Here we show that it is in fact consistent.

Theorem 2.6.1. *Suppose that $\mathbb{B}$ is a complete atomic Boolean algebra with countably many atoms. Then there is an extension of $\mathbb{B}$, $\mathbb{B}' \geq \mathbb{B}$, such that for every function $t : Atom(\mathbb{B}) \to [0, 1]$ there is a proposition in $\mathbb{B}'$, p_t, with the following property.*

> *Any countably additive probability function Cr over $\mathbb{B}$ can be extended a probability function Cr' over $\mathbb{B}'$ in a way such that:*
>
> 1. *$Cr'(p) = Cr(p)$ for each $p \in \mathbb{B}$*
> 2. *$Cr'(p_t \mid w) = t(w)$ for each $w \in Atom(\mathbb{B})$, provided $Cr(w) > 0$.*

Proof. Let $W := Atom(\mathbb{B})$. Without loss of generality we may identify $\mathbb{B}$ with $\mathcal{P}(W)$. Given $p \subseteq W \times [0, 1]$ let $p_w := \{x \mid \langle w, x\rangle \in p\}$. We define the expansion as follows

- $\mathbb{B}' := \mathcal{P}(W \times [0, 1])$
- $Cr'(p) = \sum_{w \in W} \lambda(p_w).Cr(w)$ where λ is the Lebesgue measure.
- $p_t := \bigcup_w (\{w\} \times [0, t(w)])$

$\mathbb{B} \leq \mathbb{B}'$ via the map $h : p \to p \times [0, 1]$. Note that $Cr'(p \times [0, 1]) = \sum_{w \in W} \chi_p(w).Cr(w) = \sum_{w \in p} Cr(w) = Cr(p)$, where χ_p is the characteristic function of p. This shows that 1. is satisfied.

Also $Cr'(p_t \mid w) = \frac{Cr'((\bigcup_{w \in W} \{w\} \times [0,t(w)]) \cap (\{w\} \times [0,1]))}{Cr'(\{w\} \times [0,1])} = \frac{Cr'(\{w\} \times [0,t(w)])}{Cr'(\{w\} \times [0,1])} = \frac{t(w).Cr(w)}{Cr(w)} = t(w)$. So 2. is satisfied.

□

The consistency of PLENITUDE is a simple corollary of this result. We simply identify the set of conceptually coherent priors with the set of probability measures on $\mathbb{B}'$ which extend regular probability measures over $\mathbb{B}$ and satisfy $Pr(p_t|w) = t(w)$ for each atom w of $\mathbb{B}$.

The principle of plenitude in action

If our evidence in the scenario where we have seen a tree from a distance is a vague proposition, as I have argued it is, there ought to be a proposition such that the result of updating on it results in the kind of credences you'd expect to be in; something like the credence distribution in figure 2.3. At the end of section 2.4 we argued that the principle of plenitude does exactly this: it provides us with such a proposition.

Suppose, as in section 2.4, that my prior credence over the possible heights of the tree less than 1000cm is uniform, and is given by $Cr(w_i) = \alpha$, where w_i is the proposition that the tree is $i$cm, and my posterior credence function after learning that there is a small number of peas is given by $Cr'(w_i)$. By the principle of plenitude there is a proposition p_t such that $Cr(p_t \mid w_i) = t(w_i) = Cr'(w_i)$ for each i. We'll show that p_t is the required update to get our credences from a uniform distribution to a distribution looking like figure 2.3. I.e. $Cr(\cdot \mid p_t) = Cr'(\cdot)$.

1. $Cr(w_i \mid p_t) = Cr(p_t \mid w_i).\frac{Cr(w_i)}{Cr(p_t)}$ by Bayes' theorem.
2. $Cr(w_i \mid p_t) = t(w_i).\frac{Cr(w_i)}{\sum_j Cr(p_t|w_j).Cr(w_j)}$
3. $Cr(w_i \mid p_t) = t(w_i).\frac{\alpha}{\sum_j t(w_j).\alpha}$
4. $Cr(w_i \mid p_t) = t(w_i).\frac{1}{\sum_j t(w_j)}$
5. $Cr(w_i \mid p_t) = t(w_i)$ since $\sum_j t(w_j) = \sum_j Cr'(w_j) = 1$
6. Thus $Cr(\cdot \mid p_t) = Cr'(\cdot)$ since they coincide on the worlds.

Chapter 3

Vagueness, Uncertainty and Desire

Some numbers are small, for example the number 1; some number aren't, such as 1,000,000,000. It is well known that from these premises one can show in classical logic that there is a last small number: a number N, such that N is small and $N+1$ isn't. Furthermore neither you nor I know which number that is. If you are not convinced by this latter claim ask yourself which number you think that N is: if you are unable to produce a satisfactory answer, I would suggest that this is because you do not know which number N is.[1] It therefore follows from very plausible assumptions that there is a last small number and we do not know which it is.

Some people are clearly bald, some clearly aren't and others are neither clearly bald nor clearly not bald. People of the latter sort are borderline cases of being bald. Suppose that Harry is such a person. By the law of excluded middle either Harry is bald or he isn't. Furthermore, nobody knows whether he's bald or not. Therefore either Harry is bald or he isn't, and we don't know which.

Epistemicism is notorious for making claims like this. Epistemicists argue that vague predicates have sharp cut-off points, cut-off points whose location we don't (and can't) know. Epistemicism regularly invites the 'incredulous stare'; it is thought to be paradoxical in virtue of its having this consequence. Yet haven't we just proved these consequences from very plausible premises? The law of excluded middle is prima facie much weaker than epistemicism, and indeed many non-epistemicists accept it.[2] The claim that we do not know whether Harry is bald also seems eminently plausible and is also, prima facie, weaker than epistemicism. Yet from these two claims it follows that either Harry is bald or he isn't and we do not know which. Similarly it follows from classical logic and the plausible claim that there is no number we know to be the last small number that there is a last small number and we don't know which it is.[3]

The claim that there's a last small number and we don't know which it is inevitable given classical logic and the claim that we are ignorant about the vague. If this claim is what invites the incredulous stare, then it is something not specific to epistemicism. Indeed I think that once the proofs of these claims are properly considered many of the objections usually levelled at epistemicism should be dropped on the grounds that they overgenerate. However there are some aspects of epistemicism that go beyond the commitment to

[1] Not every one agrees. See Wright [147], Dorr [24] and Barnett [6].

[2] See, for example, Fine [43], McGee & McLaughlin [92], Edgington [27], Eklund [29].

[3] I think that once we have accepted that Harry is bald or he isn't and we don't know which, it is a small step to accepting that me *cannot* know which. For the reasons we do not know whether Harry is bald have nothing to do with our impoverished epistemic status; it is not like we could go about collecting evidence until we found out whether he's bald. At any rate, it is certainly not this strengthening that makes epistemicism sound paradoxical.

classical logic and vagueness induced ignorance which are, at least, partly responsible for the incredulity. Epistemicists talk as if vague propositions are factual in a way that other classical logicians reject. Assuming some notion of 'factuality' we might put the repugnant epistemicist commitment as follows:

> There is a fact of the matter whether Harry is bald or not. (3.1)
>
> There is a last small number, and there's a fact of the matter concerning which that number is. (3.2)

Call someone who has these commitments a 'factualist'. Some philosophers have argued that this putative distinction, characterised by a commitment to (3.1) and (3.2), between epistemicism and other classical views accepting the ignorance thesis is in fact an illusion.[4] The problem for a non-factualist is to explain what these supposed commitments amount to. If the non-factualist can say nothing more about what it means for there to be no fact of the matter whether p than saying we cannot know whether p then she has not succeeded in distinguishing herself from the epistemicist.

In §3.1 of this chapter I shall consider several putative ways in which a non-factualist might want to distinguish herself from the factualist. These include disputes about whether vague predicates have sharp cut-off points, whether vagueness involves truth value gaps, whether it involves giving up on inferences such as reasoning by cases, whether vagueness involves there being multiple precisifications of the language, whether vagueness involves semantic indecision, and whether vagueness involves genuine uncertainty or ignorance. I reject all of these ways of distinguishing non-factualism from factualism about vagueness.

In §3.2 I consider what I take to be the most promising line for characterising non-factualism about vagueness: that uncertainty due to vagueness is not genuine uncertainty. This approach has been adopted in various forms by several recent authors; in this section I argue against these approaches and defend the view that uncertainty that arises from vagueness is genuine uncertainty both with regards to its functional role and in particular its role in informing our decisions.

In §3.3 it is argued that we should be probabilistically coherent even when there are vague propositions in the field of propositions over which we are uncertain. This is shown from a decision theoretic viewpoint, in which claims about the agents preferences over different actions are taken as basic. Finally in §3.4 I defend my own view in which non-factualism is partially characterised by the view that we cannot be rationally certain in p if we are certain that it's indeterminate whether p. However, the view does not collapse into a purely epistemic reading of 'it's indeterminate whether p', or 'there's no fact of the matter whether p', since I also subscribe to the principle that we shouldn't care *intrinsically* about the indeterminate. Of course, one may care whether p in virtue of the precise things that p entails; the exact statement of the thesis runs:

> If the strongest precise proposition p entails is the same as the strongest precise proposition q entails then you should be indifferent between p and q, provided p and q are maximally specific propositions.

I argue that this principle partially elucidates non-factualism and I show how factualists who deny this principle can expect to encounter concrete differences in rational decision making between people who care intrinsically about the vague.

3.1 Factualism and non-factualism

The problem outlined has traditionally been posed as a challenge for supervaluationism. On the face of it supervaluationists make exactly the same assertions as the epistemicist regarding a Sorites paradox: the challenge for the supervaluationist is to say something

[4] See Williamson [139], [140] and Dorr [24].

that distinguishes herself from the epistemicist. However, the challenge as stated above seems to be much more general; it extends to any view that accepts classical logic and the ignorance thesis.[5]

Supervaluationism has found wider acceptance in the philosophical community than epistemicism. However it is hard to explain why this is if they really are effectively the same view. What is the distinction between classical views which have seemingly unacceptable consequences, such as epistemicism, and those that do not, such as supervaluationism, if not the claims about the existence of unknowable boundaries? I would suggest the distinction is best drawn as a distinction between factualist views, of which epistemicism is just one particular kind, and non-factualist views, of which supervaluationism is just one kind. I shall briefly discuss some putative ways of drawing the factualist/non-factualist distinction taking epistemicism as the primary example of a factualist view. In each case I'll argue that either the apparent difference isn't really a difference, or that the difference isn't really getting at an important distinction.

Sharp cutoff points. Epistemicism is often is often characterised by the view that vague predicates determine sharp boundaries whose location we are ignorant of. Could it be the existence of sharp boundaries that distinguishes epistemicists from classical non-factualists?

The answer depends crucially on what we mean when we say a predicate F has a 'sharp boundary' along a Sorites sequence. It cannot simply mean that there is a last element in the sequence which is F: i.e. that there is an element in the Sorites which is F such that the next element in the Sorites sequence is not F. This much is guaranteed by classical logic and the fact that some but not all elements of the Sorites are F.

Nor could it mean that there is a clear cutoff point: i.e. an element of the sequence that is clearly F and such that its successor is clearly not F. Here both the epistemicist and the non-factualist agree that vague predicates do not have clear cutoff points; indeed there is no reason why an epistemicist and a non-factualist should disagree about first order questions about which cases are clearly F and which are borderline.

Finally, things would not look much better if we treated a sharp boundary as one whose location is entailed by the physical facts. Both factualists and non-factualists alike usually agree that whether someone is bald, for example, supervenes on more basic physical facts such as their hair number, colour and distribution. So a non-factualist, like the epistemicist, will accept the disjunction: either the physical facts strictly imply that Harry is bald or they strictly imply that he's not bald (although it's vague which.)

Truth value gaps. Many people have argued that vagueness gives rise to truth value gaps: propositions (or sentences) which are neither true nor false. Bivalence is the claim that every proposition (sentence) is either true or false and should be sharply distinguished from the law of excluded middle, the schema $p \vee \neg p$.

Does this explanation of vagueness meet the challenge of giving a non-epistemicist reading of ∇? Certainly cashing out vagueness in terms of an antecedently understood alethic notion achieves this. But the notion of truth appealed to here is not the pretheoretic one, according to which 'p and it's not true that p' is inconsistent. Surely both factualists and non-factualists believe in the existence of two operators in the vicinity, one behaving disquotationally the other not.[6] Is the difference really a terminological one of what to call truth and what to call clear truth? If it is a non-terminological dispute about how vagueness interacts with some particular role that truth plays it would be better, I think, to phrase the debate in terms of that role.

Multiple precisifications. A natural way for the supervaluationist to distinguish herself from an epistemicist is to employ the language of precisifications. As a first try: supervaluationists believe there are multiple admissible precisifications of a vague language

[5] [139], for example, makes makes this general argument.

[6] For example, in higher order logic there is a comprehension principle for operators, from which it follows that $\exists T \forall p(Tp \leftrightarrow p)$ and $\exists T \forall p(Tp \leftrightarrow \Delta p)$.

and epistemicists believe there is only one.

In order to make this precise, let us assume there is some set of admissible precisifications, V – I shan't make any assumptions about what they are.[7] For each precisification $v \in V$ we also have the notion of a formula holding at that precisification: $v \Vdash \phi$. We shall assume that for each v, $v \Vdash$ behaves like a normal modal operator and also that it commutes with the extensional connectives: $v \Vdash \neg\phi \leftrightarrow \neg v \Vdash \phi$ and $v \Vdash (\phi \vee \psi) \leftrightarrow (v \Vdash \phi \vee v \Vdash \psi)$. The constraint on precisifications is therefore that ϕ is determinate iff, for every $v \in V$, $v \Vdash \phi$.

Unfortunately it is easy to introduce things behaving like V and $\Vdash$. A natural way to do this is to let precisifications be maximally consistent propositions, i.e., consistent propositions such that for any other proposition, it entails that proposition or its negation. Say that $v \Vdash p$ only if v entails p. Finally, say v is admissible iff it's not determinately false, i.e. $\neg\Delta\neg v$ (alternatively we could say that v is admissible iff $\forall p(\Delta p \rightarrow v \Vdash p)$.)[8]

What is admissible in this setting depends on what Δ means – it could, for all we've said, have an epistemic reading. The purely formal claim that a sentence is determinate if it is true on all admissible precisifications has no content unless we've said what 'admissible' means. Informally a precisification is admissible just in case makes all the clearly red things red, all the clearly non red things not red, and so on. If this is what 'admissible' means, then even the epistemicist can say there are multiple admissible precisifacations. (On the other hand, note that both epistemicists, supervaluationists and any classical logician is committed to the existence of exactly one disquotational precisification. So the sense in which epistemicists are committed to exactly one correct interpretation applies to the supervaluationist too.)

Semantic indecision. McGee and McLaughlin, Lewis, and many others, have argued that vagueness is a matter of semantic indecision. On the surface this seems like it is in stark contrast with epistemic views, which urge that vagueness is an epistemic phenomenon. However it is unclear whether semantic indecision really amounts to something that different from the epistemic view. Its proponents often spell out semantic indecision as a failure of our use of the language to fix the truth conditions of the relevant sentences. As McGee and McLaughlin put it 'the truth conditions for a sentence are established by the thoughts and practices of the speakers of the language [...] a sentence is [definitely] true only if the non-linguistic facts determine that these conditions are met' [92].

What does it mean for a sentences truth conditions to be 'determined', 'fixed' or 'established' by the linguistic practices of the users of that sentence? A standard slogan has it that meaning supervenes on use; that any two worlds in which all the facts about the way a language L is used is the same are worlds in which the meaning facts about L are the same.[9] So it seems at though the determining relation is not to be cashed out in terms of supervenience or strict implication but in epistemic terms. This brings the view very close to Williamson's view, in which vagueness consists in the linguistic practices of L speakers not epistemically determining the truth conditions of L's sentences.

Logic. I am restricting my attention to classical views; views which accept all classical validites. However, views which are classical in this sense, while agreeing on the validities, can differ about what counts as inconsistent and can differ about what follows from what.

According to some theorists logical entailments should preserve determinate truth (as opposed to disquotational truth.) In this setting various classical rules of proof fail, such as contraposition, reasoning by cases, and reductio ad absurdum, when the Δ operator is

[7] Due to higher order vagueness, 'V' here cannot be assumed to be a precise name or a set

[8] Here I have assumed a non-linguistic indeterminacy operator. For a linguistic theorist, we could let precisifications be models of the language, and we could let the admissible precisifications be models M which make true every definite sentence. I.e. $V = \{M \mid \forall s(Def(s) \rightarrow M \models s)\}$.

[9] This slogan is false once we admit the possibility of reference magnetism (see [69].) However, since the sense in which 'electron"s meaning is not determined by use is not the kind of thing that makes 'electron' vague this seems like just another argument that we cannot treat 'determines' ('fixes' etc) as expressing a supervenience claim.

in the language.

Let us begin by stressing an important distinction between

> People who accept all classical theorems.
>
> People accept that ϕ is logically valid, for every classical theorem ϕ.

By a classical theory of vagueness I mean someone of the first type, not the second. I might have some weird and interesting views about about what logical validity is, and yet still accept all the problematic instances of excluded middle involving borderline cases. This point applies to logical consequence just as forcefully as it applies to logical validity. If the same principles govern the way in which two people accept new things from old, then an uninteresting question (for the purposes of the present discussion of vagueness) is which of the many related constraints on acceptance and credences we should call logic (e.g. are valid arguments the certainty preserving arguments, or are they the arguments that permit no increase in credence from conclusion to premise?)

Does the rejection of classical rules of proof distinguish the non-factualist reading from the epistemic reading of ∇? We seem to once again run the risk of this turning into a terminological dispute. The question is: does the factualist and the non-factualist agree about all the facts not involving consequence talk? If so, it's entirely possible that they are arguing about what we call a consequence. Let $p \vdash_1 q$ if for every rational credence function $Cr(p) \leq Cr(q)$ and let $p \vdash_2 q$ if for every rational credence function $Cr(p) = 1$ only if $Cr(q) = 1$. Note that reductio, for example, plausibly holds for $\vdash_1$ but not $\vdash_2$.[10] Both are well defined relations. Two theorists may agree about which credence functions are rational to have, yet disagree about whether reductio holds because because they are calling different things consequences. What would *really* count as a difference would be if they both disagreed about whether $\vdash_1$ obeys reductio. This would frame the dispute in terms of what kind of doxastic propositional attitudes it's rational to have towards vague propositions. This is what I take to be the most fruitful way of formulating the dispute out of the ways that we have discussed so far, and it is to this that I now turn.

3.2 Vagueness and uncertainty

On the face of it uncertainty about the truth of vague propositions seems to be of a very different kind from uncertainty of the more commonplace kind. Uncertainty about the future, about the goings on in far away galaxies, or about wildlife in the deep sea seems to have a very different nature. For one thing, the latter kind of uncertainty appears to be uncertainty about the way the world is. According to the non-factualist, the former kind cannot be like this since the world simply does not give answers to questions that have no determinate answer. Secondly, ignorance or uncertainty of the latter kind is curable; there are facts one could learn that would put one in a better position to judge whether or not p. It is difficult to even imagine what it would be like to be in a better epistemic position with regards to p, if p were vague. If I knew all there was to know about Harry's number, distribution, and so on, and I still didn't know whether he was bald, it's hard to see what else could I do to decide the question - there is nothing left to learn that is relevant to his baldness.

However many people, myself included, find this talk of 'worldly uncertainty' hopelessly obscure. In recent years a number of philosophers have situated the study of vagueness related uncertainty at the center of the puzzles surrounding the nature of non-factual propositions instead ([37], [118] [6], [24], [123], [86], [134].) The operative thought, according to many of these philosophers, might be encoded in the following way:

[10] Reason: it seems independently plausible that for no rational credence function $Cr(p \wedge \neg\Delta p) = 1$, so $p \wedge \neg\Delta p \vdash_2$, yet surely if we're certain p is indeterminate then $Cr(\Delta p) = 0$ and $0 < Cr(p) < 1$ so $Cr(\neg(p \wedge \neg\Delta p)) < 1$. So $\nvdash_2 \neg(p \wedge \neg\Delta p)$

NF. One may be genuinely uncertain whether p only if one considers there to be a fact of the matter whether p.

If one accepts this principle then we have a fairly natural way of explaining the difference between factualism and non-factualism about a subject matter.

In order to be a non-factualist about vagueness, then, we need some explanation of what our doxastic state looks like when we do not consider there to be a fact of the matter whether p. Do we fail to be uncertain, or are we just not 'genuinely' uncertain? I split my discussion of this up into three different approaches to this question:

> **Ersatzism**. When one is uncertain due to vagueness one is in a *sui generis* mental state, a state that cannot be explained in terms of the kind of psychological vocabulary we use to describe ordinary uncertainty.
>
> **Rejectionism**. When one has accepted that it's vague whether p, one should reject p and reject $\neg p$. Similarly, one's credences should fail to be additive in accordance with the rule of assigning the same credence to p as to the claim that it's determinate that p.
>
> **No ignorance**. The source of one's uncertainty is never due to vagueness; if one knows the relevant facts about Harry's hair number, distribution, and so on, one also knows whether Harry's bald.

In what follows I shall argue that none of these proposals work and that one should therefore reject the principle NF. If you know that there's no fact of the matter whether p you may, and indeed often should, be genuinely uncertain whether p. This calls for another explanation of the difference between factualism and non-factualism, to which I turn in section 3.4.

3.2.1 Ersatz uncertainty

According to ersatzism when one is in the kind of doxastic state characteristic of being uncertain in a proposition you know to be vague, you are in fact not uncertain in the ordinary sense, but in some other *sui generis* mental state. Talk of 'uncertainty' in cases of vagueness relies on an equivocation between two propositional attitudes. For example Schiffer - a prominent defender of this kind of view - claims that vagueness-related uncertainty is 'a new kind of propositional attitude, one that comes in degrees and that precludes standard partial beliefs,' that it is '*not* a measure of uncertainty', and similar such things. I shall call this state 'ersatz uncertainty'. If such uncertainty comes in degrees, we shall also talk about 'ersatz credences'. (See [118], [123] and [86] for different ways of formulating the view.[11])

Ersatz uncertainty, if it were to exist, would give rise to 'mixture' uncertainties. I give two examples:

> You are genuinely uncertain whether p is vague or precise. Your credence in p should be a mixture credence: perhaps this is your conditional credence in p given p is vague (an ersatz credence) times your genuine credence that p is vague, plus your conditional credence in p on p being precise (a genuine credence) times your genuine credence that p is precise.
>
> p is second order vague. You are ersatz uncertain whether p is vague or precise. Your credence in p should be a mixture credence: perhaps your conditional credence in p

[11] Note that not all of these elaborations on Schiffer's idea count as genuine ersatz theories on my account. For example Smith, in arguing that degrees of belief are expected truth values, accepts a view in which there is really only one kind of doxastic state, albeit one that does not obey the classical probability calculus. We consider views like this in the next section.

given p is vague (an ersatz credence) times your ersatz credence that p is vague, plus your conditional credence in p on p being precise (a genuine credence) times your ersatz credence that p is precise.

In order to see that there must be mixed uncertainty of the first kind suppose you have has a bag containing two small coloured balls. The first is clearly red, the other bordering on red/orange. You are about to take a ball out of the bag: what should your credence that you will pick a red ball be? This is a mixture uncertainty of the first kind: you are genuinely uncertain whether the ball you pick will be the first or second ball, yet given it's the second ball you should be ersatz uncertain whether it's red. A standard way to approach this question would tell you to work out the probability that it's red given that it's the first ball and the probability that it's red given it's the second ball, then times each of these by the probability that it is in fact the first/second ball respectively, and add these together: $Cr(red) = Cr(red|ball1) \times Cr(ball1) + Cr(red|ball2) \times Cr(ball2)$. Since in this instance $Cr(red|ball1)$ represents the numerical value of a genuine credence and $Cr(red|ball2)$ the numerical value of your an ersatz credence, $Cr(red)$ is a mixture credence.

In order to see that there must be mixed uncertainty of the latter kind imagine instead that you have lined up in front you a sequence balls starting with a clearly red ball and ending with a clearly orange ball. This constitutes a Sorites sequence for redness. Presumably according to the ersatz theorist we should have high genuine credences that the balls at the beginning of the sequence are red, intermediate ersatz credences that the balls in the middle are red, and low genuine credences that the balls at the end are red.[12] One might reasonably wonder what happens at the transitions between the genuine and ersatz credences. Surely we don't know where these are due to vagueness of where *these* boundaries lie (just as we don't know where the boundary between red and non-red lies.) In these situations we are ersatz uncertain whether the balls between the beginning and the middle of the sequence are determinately or vaguely red. Let b be a ball such that it's vague whether it's determinately red or not, and let B be the proposition that b is red. So it seems that typically people in this situation will have a mixture credence, this time given by $Cr(B) = Cr(B|\Delta B).Cr(\Delta B) + Cr(B \mid \neg\Delta B).Cr(\neg\Delta B)$, where here $Cr(\Delta B)$ and $Cr(\neg\Delta B)$ are ersatz credences.

It seems that the best way to account for the two kinds of situations described above is to assign an important role to a kind of mental state which is partly determined by your genuine credences and partly determined by your ersatz credences. For lack of a better name, call these mixture credences 'degrees of belief' (following Smith [123].)[13] Now some degrees of belief are pure in the sense that either the coefficient of the genuine or ersatz part is 0; we might adopt a convention of calling the pure genuine degrees of belief genuine credences and everything else an ersatz credence, or vice versa. But it seems natural to think that really there is just one doxastic attitude at work in all cases – one's degree of belief.

There is a worry that this issue is terminological. The initial target was a person who thought vagueness didn't *really* involve uncertainty, and talk about uncertainty in such cases would better be attributed to some other attitude. One might accuse me of using the term 'genuine uncertainty' more elastically than my opponents - after all, I certainly admit there's something distinctive about vagueness-related uncertainty. Perhaps this is so; and if it is, I have at best clarified how much mileage one can get from the claim that vagueness doesn't involve genuine uncertainty.

However, the position was introduced as one in which, in order to describe a persons vagueness related beliefs, we need to introduce new psychological vocabulary which is not reducible to the vocabulary we ordinarily use to describe their doxastic attitudes. But the

[12] It is interesting to note that if this is the picture there is no such thing as ersatz certainty; whenever you have a numerical credence corresponding to a ersatz state of uncertainty, it will have an intermediate value.

[13] For discussion about how this might work see [85], [86] and [123].

concession that we have degrees of belief as well as genuine and ersatz credences poses a serious problem to this view. The vocabulary we *ordinarily* use to describe a persons doxastic state will almost always be a degree of belief – a mixture of ersatz and genuine credences – and almost never be a pure genuine or pure ersatz credence since it rarely, if ever, happens that an agent can rule out all the situations in which a proposition would be borderline. The view was supposed to entail that the uncertainty we have about whether Harry is bald is fundamentally different from the uncertainty we have about whether there is non-human intelligent life in the Milky Way. Yet in reality both these examples involve mixture uncertainties: I may have a small credence that Harry is wearing a bald head wig, and actually has a full head of hair, and I may have a small credence that the only non-human life elsewhere in the galaxy is however intelligent it needs to be to be borderline intelligent.

I think even if the debate is partially terminological, we *do* have a reasonably good account of when something counts as genuine uncertainty. For example, you do not need to be human to be genuinely uncertain about something. Presumably animals with a very different neurophysiology from us could still be genuinely uncertain. We can recognise this, and there is no terminological question about whether they count as genuinely uncertain. The brain state I am in when I'm genuinely uncertain about something could be radically different in configuration and constitution from the same state of genuine uncertainty in an alien. It is not the actual brain state, but the causal-functional role it plays with respect to the rest of the being that makes it a state of genuine uncertainty.

It could be that, physiologically, vagueness-related uncertainty is very different from ordinary uncertainty. But as with the case of the alien, this does not mean that vagueness-related uncertainty is not genuine uncertainty. To establish that vagueness-related uncertainty is not really genuine uncertainty, one must show that it plays a very different causal-functional role. On first looks, however, this does not seem to be the case. Someone who is in a state of vagueness-related uncertainty in p because they're certain p is vague, typically won't assert p or $\neg p$, won't stake things they desire on p, and will generally act as if they where uncertain about p. If degrees of belief are playing the same causal-functional role as genuine uncertainty is supposed to be playing with respect to our actions and desires, as is suggested above, there is little substantial left to the claim that it is not genuine uncertainty. It is insubstantial in the same way in which it is insubstantial to insist an alien with a different physiology can't be genuinely uncertain – it is a verbal disagreement at best, since we do not need to introduce new psychological vocabulary to describe the alien's psychological state.

3.2.2 Rejectionism

Let us assume, then, that degrees of belief form a single psychological natural kind. One might then conclude that, since the difference between vagueness related beliefs and ordinary beliefs is not one in kind, it is a difference in the normative principles governing them. Schiffer argues that ersatz credences are governed by the rules for calculating the truth values of propositions according to the Łukasiewicz connectives; one way to take on board the message of the last subsection is to take the notion of a degree of belief, i.e. a mixture credence, as the operative psychological concept, but identify situations in which one's degree of belief behave non-classically as being the distinctive feature to do with vagueness related beliefs.

One such way of doing this within the Schifferian framework is to identify one's degree of belief in a proposition with its expected truth value valuated according to the Łukasiewicz connectives (see Smith [123].) Unfortunately this makes use of the ideology of degree of truth; we have dealt already, for classical views at least, with the view that vagueness is to be identified with a further truth status. Furthermore, according to this proposal if you're certain that p has degree of truth $\frac{1}{2}$ then $Cr(p \wedge \neg p) = \frac{1}{2}$, yet a state assigning this value

to a contradiction couldn't play the right causal functional role of a credence, which would in part involve telling us to reject bets on contradictions.

There are a number of discussions of how classical probability should be weakened to deal with vagueness related beliefs (see e.g. [37], [23], [123],[86]); my focus will be mainly on the proposal of Field in [37].[14] Field's theory is rich and technically involved, but the full details are not important for what I say here. An important consequence of the theory is that your credence in p and in Δp should always be the same. One characteristic of such a theory is that if there are any rational credence functions assigning non-zero probability to the claim that it's indeterminate whether p they will be non-additive. In particular:

For any rational credence function Cr, if $Cr(\nabla p) = 1$ then $Cr(p) = Cr(\neg p) = 0$.

This is clearly incompatible with finite additivity, which requires the sum of the probabilities of a proposition and its negation to add up to 1. Following Williams [136] I shall call the general approach 'rejectionism': when you're certain p is vague you should reject both p and its negation. In Field's framework rejection is given a concrete interpretation in terms of having no credence in p. The crucial thing about this proposal is we are not postulating a fundamentally different type of psychological state, we are simply revising the laws of probability for the ordinary notion of credence.

However Field's proposal has a number of difficulties, which I'll address in order of seriousness.

Problem 1: Field's calculus delivers the wrong intuitive predictions. According to Field if one is certain p is vague, then one's credence in p should be 0. Yet there are cases where you're certain that p is vague, but have an intermediate, or even a very high credence in p. Imagine I'm about to roll a hundred-sided die, whose sides are labelled 0 to 99. The probability that the die will land on any particular number is surely 1%. So intuitively I should be fairly sure that the number it's going to land on won't be the last two digits of the age in nano-seconds at which I stopped being a child; which is, after all, a particular number. There's only one of this number, and 99 of the others, and whichever one it is, there's a 99% chance it won't land on that one due to the fact that the die is fair. So I should be 99% certain that it's not going to land on the last two digits of time in nanoseconds at which I stopped being a child. But I'm absolutely certain that whichever number it does land on, it'll be vague whether it's the last two digits of the critical time at which I stopped being a child, because I'm certain there are at least 100 borderline cases (I was a borderline child for well over a second, so here will be thousands of borderline cases.) Prima facie, this seems to be a case in which $Cr(\nabla p) = 1$ and $Cr(p) = .99$, contradicting Field's prediction that $Cr(p) = 0$.

Problem 2: According to Field if you are certain that p and q are vague, you cannot think that p is more likely than q, or that q is more likely than p. They both receive credence 0. But this, once again, does not accord with our intuitive judgements. Suppose you are looking a two glassing, on of which is 65% full the other 67% full. Now you might be certain that it's vague, in both cases, whether the glass is pretty full, but you should be *more* confident that the second is pretty full, since it is clearly more full than the former. Indeed, once we can make sense of comparative judgements of confidence in the face of vagueness it is but a short step to proving we must be probabilistically coherent via Villegas' theorem (i.e. that, given certain plausible conditions on the ordering of comparitive judgements of confidence, one can uniquely represent confidence by a probability measure.)

Problem 3: Field assumes a logic of S4 for Δ, and this assumption is central in his construction of a probability function satisfying his formal requirements. In fact, S4 is forced on him in a certain sense since his theory entails that for every rational credence function $Q(\Delta\Delta p) = Q(\Delta p)$ and (less obviously) that $Q(\neg\Delta\Delta p) = Q(\neg\Delta p)$ showing that Δp and $\Delta\Delta p$ are probabilistically equivalent. Another reason Field needs S4 is because he

[14] Field later develops the view in a setting suitable for a non-classical logic; I shall focus exclusively on the classical version of the proposal here.

wants his theory of rational credences to uniquely single out Δ as playing the role of the definitely operator. It is easy to see, however, that if an operator O satisfies his constraints (such as $Q(Op) = Q(p)$) so would the operator OO, OOO, and so on; S4 would identify all these iterations.

The problem is, S4 is not a suitable logic for dealing with higher order vagueness. S4 does not *obviously* rule out higher order vagueness; it's not obvious that there can't be propositions p, such that it's indeterminate whether p is determinate or not, although by S4 p cannot be determinately true in these cases. As Field writes

> "It is sometimes thought that the S4 law is inappropriate in connection with higher order vagueness. That is certainly not obvious: indeed, we have seen above that significant higher order vagueness can be recognized [...] as long as S5 is not assumed"

However it is possible to show that the following 'gap principle' entails that everyone is bald in S4. Let p_i be the proposition that someone with i hairs is bald. Then we get a contradiction from the assumption that Δp_0, $\Delta\neg\Delta p_{10000}$ and the following gap principle:[15]

$$\Delta(\Delta\Delta p_i \rightarrow \neg\Delta\neg\Delta p_{i+1})$$

Denial of this principle implies there is a sharp, determinate, cut off between the determinately bald and the not determinately bald. So denying this principle effectively amounts to saying that there is no second order vagueness in a Sorites sequence of this form (analogous points apply to orders higher than two.)

Problem 4. Suppose that I know that it's vague whether Harry is bald. Standard decision theory says that the expected utility of an action is identical to the expected utility of the action given Harry is bald times the probability that Harry is bald plus the expected utility of the action given Harry is not bald times the probability that Harry is not bald. However according to Field this result would yield an expected utility of 0 whatever the action was since we are multiplying by each summand by 0. Since we are supposed to be always be able to calculate the expected utility of an action relative to the most fine grained partition (in which every element is *a priori* vague) this result would be devastating for Field. We must therefore consider ways to revise the standard decision theory.

Let Q be a Fieldian credence function, and P a probability function (so P obeys the Kolgomorov axioms.) Say that Q admits P just in case $Q(p) \leq P(p)$ for every proposition p. Let $A(Q)$ denote the set of probabilities Q admits. Given a probability function P and given a real valued random variable u, write $E_P(u \mid p)$ for the conditional expectation of u given p relative to P. Then we can then define a set valued notion of expected value: Let $Val_Q(u \mid p)$ be the the set $\{E_P(u \mid p) \mid P \in A(Q)\}$. Here are two principles governing how our preferences should be determined from our Fieldian credences and our utilities u.

$p \preceq q$ if and only if $E_P(u \mid p) \leq E_{P'}(u \mid q)$ forall $P, P' \in A(Q)$

$p \preceq q$ if and only if $E_P(u \mid p) \leq E_P(u \mid q)$ forall $P \in A(Q)$

$p \preceq q$ if and only if $\inf Val_Q(u \mid p) \leq \inf Val_Q(u \mid q)$

Note that in this framework the objects of your preferences are not actions but propositions. To say that I prefer p to q (written above as $q \preceq p$) means, roughly, that my happiness conditional on learning p is greater than my happiness would be conditional on learning q.

The first proposal does not seem to accord with our intuitions well. For example, let us suppose that I'm presented with a bag containing 4 balls. I know that 3 of them are clearly

[15] Let $\mathcal{M}p := \neg\Delta\neg p$. Note that $\Delta(p \rightarrow q) \rightarrow (\mathcal{M}^n p \rightarrow \mathcal{M}^n q)$ is a theorem of S4. So the gap principle entails $\mathcal{M}^n\Delta\Delta^m p_i \rightarrow \mathcal{M}^n\mathcal{M}\Delta^m p_{i+1}$. Then we reason with the following chain: $\Delta p_0 \rightarrow \Delta^{10000} p_0 \rightarrow \mathcal{M}\Delta^{9999} p_1 \rightarrow \mathcal{M}^2\Delta^{9998} p_2 \rightarrow \ldots \rightarrow \mathcal{M}^{9999}\Delta p_{10000}$. So $\mathcal{M}^{9999}\Delta p_{10000}$. But since $\Delta\neg\Delta p_{10000}$, we get by S4 $\Delta^{9999}\neg\Delta p_{10000}$, contradiction. This argument is adapted from Fara [33].

red, but one of them is a borderline red/orange ball. A ball has been selected at random; I'm pretty, but not totally, sure that the ball is red. Letting my Fieldian credence be Q and letting R be the proposition that the ball is red then $Q(R) = \frac{3}{4}$ and $Q(\neg R) = 0$. Now I'm offered two bets: one pays out £1 if the ball is red, the other pays out £1 and 20 pence. It seems obvious that I should prefer the second bet to the first, and that this should be born out in my preferences over the proposition that I accept the first bet over the secend. Writing B_1 and B_2 for the proposition that I accept the first and second bet respectively it seems as if $B_1 \prec B_2$. However this is not the case: the expected utilities of the former bet spans $[0.75, 1]$ while the latter spans $[0.9, 1.2]$.

The second proposal does better than the first with regards to the second example. However it still postulates incomparable preferences: if for some $P \in A(Q), E_P(u \mid p) < E_P(u \mid q)$ and for other $P \in A(Q), E_P(u \mid q) < E_P(u \mid p)$ then it's neither the case that p is preferred to q or q is preferred to p, or that you are indifferent between p and q. However it is hard to make sense of incomparable preferences; how might these preferences manifest themselves in your dispositions, for example? One might argue that it means that were you offered a choice between making p the case or q the case, it's not the case that you would choose p and it's not the case that you would choose q (even though you would choose p or q.) This response relies on the failure of a hotly contested principle in the logic of counterfactuals known as 'conditional excluded middle' (henceforth CEM; see Stalnaker [126] for further discussion.[16]) Furthermore, whenever the antecedent is true, i.e. you are confronted with the choice between p and q, CEM holds. Yet this threatens to force your actual decision making patterns to behave as if you had a single member of $A(Q)$ as your credence. If so, it is more natural to take your credences to be this single probability than the Fieldian function Q.[17]

The third proposal ensures that your preferences are linear at the cost of making your preferences concerning vague propositions unreasonably conservative.[18] The first casualty is a principle known as 'averaging', an instance of which is stated: if p and q are incompatible and and $p \approx q$ then $p \approx (p \vee q) \approx q$ (another version, in which $\approx$ is replaced by $\prec$, also fails.) Suppose that I am certain that it's vague whether Harry is bald. Then according to the third proposal the infimum of the expected values of $(H \wedge L)$ and $(\neg H \wedge L)$ are both 0, so $(H \wedge L) \approx (\neg H \wedge L) (\approx \top)$. However $(H \wedge L) \vee (\neg H \wedge L) = L$, which we may suppose I clearly care about; this contradicts the averaging principles which say that $(H \wedge L) \approx L (\approx \top)$.

The failure of this theoretical principle is not in itself a problem. Perhaps the possibility of beliefs about the vague casts doubt on the averaging principle. However the proposal predicts very strange preferences. For example, if I am certain that it's vague whether Harry is bald, the proposal predicts that my happiness conditional on me finding out that I have won ten pounds is greater than my happiness conditional on me finding out both that I have won a million pounds and that Harry is bald.

3.2.3 No uncertainty

Finally there are a class of views which deny that we are uncertain or ignorant about vague matters altogether. For example, Dorr [24] argues that when we know the relevant facts about how much of a particular glass is filled with water, say, we typically know whether the glass is pretty full or not. This applies even if it's a borderline case. According to Dorr, vagueness is a matter of semantic indecision. The relevant fact about vagueness in the vicinity is the fact that the sentence 'the glass is pretty full' is vague in English (at context c, time t, and so on.) This, Dorr contends, and I agree, is a fact about English and 'the claim that human beings are doomed to ignorance as regards whether the glass is pretty full [...] has nothing at all to do with language: if it is true, it would have been true

[16] We are actually relying on a more general principle, which obtains in Stalnaker's logic, namely: $\phi \rightarrow (\psi \vee \chi) \vdash (\phi \rightarrow \psi) \vee (\phi \rightarrow \chi)$.

[17] See Elga [30] for further discussion of these problems.

[18] There is a dual proposal which substitutes the 'inf' for a 'sup'. Parallel problems arise for this proposal.

no matter what we had meant by the words 'the glass is pretty full'.' If the no-ignorance view was plausible then it would render most of the arguments of chapter 1 moot.

For those who balk at epistemicist accounts of vagueness this alternative is hardly better. According to this view, not only is there an exact percentage, $\%x$ say, such that the glass is pretty full if filled to be $>\%x$ full, and is not pretty full otherwise, but we are furthermore supposed to *know* what that critical percentage x is. If we are watching a glass being filled, drop by drop, then provided we are keeping count, not only will there be a drop which makes the glass pretty full, but we'll know which that drop is.[19] The claim that we do not know at what point the glass becomes pretty full is surely not the unintuitive claim that causes people to balk at epistemicism.

Another version of the no-ignorance view is defended in Barnett [6]. Unlike Dorr, however, Barnett is not a linguistic theorist. He treats indeterminacy as an operator and presumably therefore admits vague propositions. Since, at least for these purposes, Dorr is an intensionalist, it is easy on that view to communicate whether Harry is bald in English; one of the sentences of the form 'Harry has less than n hairs' will do. If N is the critical number then the proposition that Harry is bald simply *is* the proposition that Harry has less than N hairs and both are expressed by the sentence 'Harry has less than N hairs.' We cannot, according to Dorr, communicate this fact using the sentence 'Harry is bald', this is what explains our reluctance to assert this sentence when Harry is a borderline case. For Barnett on the other hand, the proposition that Harry is bald is distinct from the proposition that Harry has less than N hairs, and the former is presumably expressed by the sentence 'Harry is bald', and not by 'Harry has less than N hairs' (which expresses the latter.) The belief that Harry is bald can be communicated with the sentence 'Harry is bald'.

It seems, therefore, as if Barnett has a special problem of explaining why we do not assert that Harry is bald, or that he isn't, when it's vague whether he's bald. Barnett appeals to the principle:

> One should aim to clearly satisfy the rule: (M) assert p only if p.

according to Barnett there will be cases where you know whether p, but you are not allowed to assert p (or $\neg p$) because it would result in its being unclear whether you've satisfied (M). That is, *even though you know you've satisfied (M)*, i.e. even though you know you've asserted p only if p, you would not have clearly satisfied (M) and this in itself is supposed to be a bad thing. No explanation for this rule is given; it is just a primitive norm of assertion.

Now, if we are to believe the claims made in [6], Barnett knows which the last small number is (providing, of course, that his cognitive conditions are ideal, and so on and so forth. See p28 of [6].) Yet due to this primitive norm of assertion, he's not allowed to tell us which it is. But even so, one might think these norms can sometimes be trumped; maybe by other norms of communication, or by moral norms or by something else. As an analogy: I may know that Jones has a certain medical condition, but even though I *know* this I shouldn't assert that he does as it would cause him great embarrassment. However, this barrier on assertion can be trumped, for example, if he were suddenly taken to hospital and the staff needed to know about his medical history.

If Barnett really does know which the last small number is it would be of great value to the rest of the philosophical community if he were to tell us. After all, there has been a long-standing debate amongst classical and non-classical logicians about whether there even is one; if Barnett could settle the matter constructively that would be significant progress. Therefore, given the great benefits, it seems like it would easily be worth the cost of flouting

[19] I take it that the 'knowing which' facts supervene on the 'knowing that' facts. For example, if you know that after drops 0-N the glass is not pretty full, and you also know that the glass is pretty full at the Nth drop and onwards, then by some closure principles it follows that you know that $N+1$ is the first drop at which the glass is pretty full, and therefore, I claim, you know which the first drop the glass is pretty full at is.

the rule and telling us which the last small number. Yet he has not told us. I think it would be not unfair to conclude that this is not because he is afraid of violating the rule (M), but because he simply doesn't know which number it is.[20]

3.3 Vagueness and desire

In this chapter and the previous chapter we have considered the interaction between the vagueness of a proposition, p, and the rationality of various doxastic attitudes you could have towards p. On the current account of this interaction when you know that it's vague whether p you should be genuinely uncertain whether p. In the rest of this section we shall consider the relation between vagueness and bouletic attitudes such as desires and preferences. We'll begin by outlining a theory of preferences in which we can state some of the decision puzzles mentioned in the last section. We shall use it to both vindicate the view that one should be uncertain when you're certain that it's vague whether p, and that one's degrees of belief should be probabilistically coherent over the space of vague and precise propositions. In §3.4 we'll then see whether within our theory of preferences can say something to prevent the classical theory of vagueness defended here collapsing into epistemicism (or any other form of factualism.)

3.3.1 Preferences

We shall shall primarily be concerned with the properties of a primitive binary propositional attitude of preference between propositions, which I shall write: $p \prec q$.[21] Roughly, we can read $p \prec q$ as saying that things would be better for the agent on the assumption that q than on the assumption that p. We shall similarly assume a parallel notion of indifference, written $p \approx q$, saying that things would be just as good for the agent on the assumption that p as on the assumption that q. Weak preference is defined $p \preceq q := p \prec q \vee p \approx q$. We shall also make use of a notion of conditional preference, written $p \prec_r q$, stating that if the agent were to add r to her stock of knowledge she would prefer q over p. Conditional preferences may be defined from unconditional preferences as follows: $p \prec_r q$ iff $p \wedge r \prec q \wedge r$, and $p \preceq_r q$ iff $p \wedge r \preceq q \wedge r$.

I shall not yet assume anything about the connection between preferences and decision theory. The statement $p \prec q$ merely states that the agent holds the attitude of preferring to the propositions p and q, that things are better for the agent given that q than that p, and says nothing about how this should relate to her dispositions to act if she were rational. The preference relation holds between propositions which we have little to no control over, so that differences in preference can obtain without having any effect on our dispositions to act. One might also think that in stating a theory of rational action one must introduce a causal element to deal with Newcomb problems [97]; this does not mean that the binary propositional attitude being studied here is inadequate or that it must be revised to account for causal considerations.

Unlike some decision theorists I shall not assume that preference is to be analysed in terms of other more basic propositional attitudes. In particular I shall not assume that preference can be reduced to a purely doxastic attitude, 'credence', and a purely bouletic attitude, 'utility', which in both cases relate propositions to real numbers. While it will eventually follow that there will be some probability function and a real valued function u such that the conditional expectation of u given p is less than the conditional expectation of u given q iff $p \prec q$, the order of priority may indeed go in the other direction. Furthermore, except in special circumstances, there will typically be many credence/utility

[20] Barnett might perhaps play down the value of knowledge to the philosophical community when it is not clear knowledge. Yet I for one would value knowledge of the last small number, clear or not, so there is certainly reason to break the norm (M) if only to satisfy my own idiosyncratic preferences.

[21] The agent whose preferences are in consideration shall be assumed to be given by the context.

function pairs which serve equally well in the above representation of preferences (although this indeterminacy disappears once we have stipulated that the credence function represents the agents confidence ranking – see Joyce [65] Theorem 4.5.) Everything I say is compatible both with there being a unique credence function which represents the agents fine grained doxastic attitudes which does not supervene on the agents preferences, or with there being no such function.

One reason to study preference as a primitive rather than a defined connective, other than to preserve the order of explanation, is that one can make some progress in evaluating the structure of preferences without making substantial assumptions about the structure of purely doxastic states such as credences. As we saw in Field's framework, the preference connective either delivers the wrong verdicts entirely or is not reducible, and not even coextensive, with a notion defined out of purely doxastic and purely bouletic notions.

A useful analogy might be drawn here between the issue of uncertainty and vagueness and the literature on ignorance and uncertainty in the many worlds interpretation of quantum mechanics. In that setting time has a branching structure so that when, for example, you flip a coin it seems, at least on the face of it, that you shouldn't be uncertain about the outcome: you know that there will be a branch in which the coin lands heads and a branch in which the coin lands tails. In this kind of context it is unclear whether any sense can be made out of uncertainty about the future (or, if some sense *can* be made of it, whether the uncertainty is 'genuine' uncertainty.) However in [21] David Deutsch noticed that even if it's unclear how to think about uncertainty and probability in a setting where time branches in this way, we still have a firm grasp of what it means for someone to prefer one option over another. It seems, therefore, that one may still talk about an agents preferences among propositions without assuming that the agent has genuine doxastic attitudes in propositions about the future; and decision theory remains, by and large, untouched by this feature.[22]

3.3.2 Reasoning about preferences

With the informal notions of preference and indifference in place we can begin to state some general principles which ought to hold of preference. These principles are intended to be intuitively motivated in a way that does not require one to make assumptions about the structure of credences. The field of the preference relation, and the principles we consider here, will throughout be assumed to apply to all propositions, both vague and precise. The choice to include vague propositions in the preference relation should, I think, not be controversial. Open any texbook on decision theory and you will see examples of decision problems, outcomes and choices, and rarely, if ever, will they be described in completely precise language.

The first and the most obvious principle states that the preference relation should be transitive: if the agent prefers p to q and prefers q to r, she ought to prefer p to r. This principle should be immediately intuitive and not at all dependent on the absence or presence of vague propositions in the ordering. For example, if I'd prefer to be rich than to be of modest means, and I'd prefer to be of modest means than to be poor then I'd prefer to be rich than to be poor. Similarly obvious is the claim that preference should be asymmetric, weak preference should be both transitive and reflexive, and that indifference should be an equivalence relation.[23] Clearly no proposition can be preferred over itself, p can only be as desirable as q if q is as desirable as p, and so on.

Notice that in the above example for any two of the three propositions – I'm rich, I'm poor or I'm of modest means – one is preferred to the other or they are equally desirable (in this example the latter happens only if they are identical.) One might think that this

[22] See [133], [51]. Greaves [51], for example, defends the view that we can make sense of preferences between various actions even though there is no genuine uncertainty in an world whose time has a branching structure.

[23] Depending on how preference, weak preference and indifference are defined and which are treated as primitive some of these principles are redundant.

generalises to any pair of propositions: either one is preferred to the other or they are equally desirable. This principle may be objectionable for reasons not having anything to do with vagueness. For example, suppose I do not have enough money for a three course meal, so I must choose between having a starter and a desert. In this case it seems reasonable for me to neither be indifferent, nor to have a preference one way or another; after all, they're two very different kinds of things – one is savoury, one is sweet.[24]

It's worth noting that in a situation like the above one a rational agent will either choose a starter or a desert (because choosing neither would be clearly suboptimal.) Isn't this fact sufficient for saying that the agent prefers one over the other? One could even strengthen this to the claim that the agent is *disposed* to choose one over the other, since according to some counterfactual logics the claim that the agent would either choose a starter or desert were she given the choice entails that either she would chose a starter given the choice or she would choose a desert given the choice.[25] For now I shall put this issue to one side; if the preference relation is not complete we may think of it as an idealisation for the purposes of reasoning about vagueness. In the context of vagueness, the principle is not neutral – it rules out the first and second definitions of preference in Field's framework, for example.

Another principle which also surely holds regardless of whether we are considering vague propositions is what is sometimes called the 'sure thing' principle or the dominance principle. These principles say, roughly, that if p is conditionally preferred to q relative to each member of a certain kind of partition then p is unconditionally preferred to q. For example, a finite version of the sure thing principle says: if e and f are incompatible, then if $p \prec_e q$ and $p \prec_f q$, then $p \prec_{(e \vee f)} q$.[26]

The unrestricted sure thing principle is well known to lead to problems arising from the dependence of states and acts on one another (see Jeffrey [62].[27]) We therefore restrict to partitions that are probabilistically independent of the acts. For example, suppose a dice is to be thrown, and I have a choice between placing a bet which pays out n if the dice lands on an n, and a bet which pays out $n+1$ if the dice lands on an n. Dominance applies (since my accepting the bet is independent of how the dice lands) and says that since, for each n, I'd rather have the second bet than the first on the assumption that the dice landed n, I should in fact prefer the second bet to the first. A similar principle can be formulated for indifference: if I am indifferent between p and q conditional on every state s, I should be indifferent between p and q simpliciter.

Unfortunately probabilistic independence is not something that can be stated using preferences alone. Bradley ([13] p20) shows that a given preference relation can be represented by two probability/utility function pairs, one of which treats A and B as probabilistically independent and another which doesn't. It could be that there is a more general notion of preferential independence which is weaker than probabilistic independence: e.g. being such that the sure thing principle holds.[28] But I think this is enough to cast doubt on the possibility of a substantive sure thing/dominance principle being stated in terms of

[24]Note that we can distinguish between not having a preference one way or the other and being indifferent: for example it's possible that I might prefer to have three scoops of ice cream over having two scoops, but still have no preference either way between having two scoops or three scoops and having a starter. This situation is impossible if 'no preference either way' is replaced with 'indifference': if you're indifferent between two scoops of ice cream and a starter and you prefer three scoops over two scoops, you're not indifferent between having three scoops and a starter.

[25]Assuming a direct relation between preferences and dispositions to act, this argument would generalise to rule out any indifferences between distinct propositions. Perhaps this is something we should also accept.

[26]A more general version says that for *any* partition, X, of r, which is pairwise probabilistically independent of p and q, if p is conditionally preferred to q relative to each member of X then p is preferred to q conditional on r.

[27]For example, one might reason fatalistically: either I'll get lung cancer or I won't (and not both.) Conditional on me getting lung cancer I'd rather smoke than not and conditional on me not getting lung cancer I'd rather smoke than not; therefore I should prefer smoking to not smoking.

[28]Note however that: $\frac{des(A \wedge E)}{des(\neg A \wedge E)} = \frac{des(A)}{des(\neg A)}$ if and only if $Pr(A \mid E) = Pr(A)$. You could therefore define this notion using a the three place notion $Epqr$ which holds iff $\frac{des(p \wedge r)}{des(p)} = \frac{des(q \wedge r)}{des(q)}$.

preferences.

Jeffrey has something vaguely representing the sure thing principle which doesn't require any restriction on the states such as probabilistic independence. This principle says that if e and f partition p, then if $f \prec e$, $e \prec p \prec f$. A proposition cannot be more desirable than its most desirable outcome or less desirable then its least desirable outcome. Call this 'averaging'.

Averaging also holds its appeal even in the context of vagueness. Continuing with our current example, we have argued that in the above situation that I'd rather find out I was rich than that I was poor. Averaging says that I'd rather find out that I was rich or poor than that I was poor and that I would rather find out that I was rich than that I was rich or poor.

The penultimate axiom of Jeffrey-Bolker decision theory is less obvious than the other principles discussed so far. The principle is a test for whether two propositions are equiprobable. While two propositions, p and q, can be ranked together in the preference order even if they are not considered equally likely to the agent, if there is a proposition r incompatible with both p and q such that $p \not\approx r$, then $p \vee r$ and $q \vee r$ cannot be also be ranked together unless p and q are actually equiprobable. r, so to speak, pulls p and q apart.

Say that a proposition r is a test proposition for p and q iff it is incompatible with both p and q, and $p \not\approx r$. The principle in question states that if $p \approx q$, r is a test proposition for p and q, and $p \vee r \approx q \vee r$, then $p \vee r \approx q \vee r$ for every test proposition r for p and q.

The final axiom effectively states that the preference relation is continuous in the sense that whenever the supremum or infimum of a set of propositions lies in a certain interval (let us say, lies between p and q) then some member of the set lies in that interval.

Putting these together we get the axioms for Jeffrey-Bolker decision theory. Since we shall appeal to them at various points, I shall list them, in the formal notation, below.

1. $\prec$ and $\preceq$ are transitive total relations. $\prec$ is irreflexive and $\preceq$ is reflexive.
   (a) If $p \prec q$ and $q \prec r$ then $p \prec r$ (and similarly for $\prec$.)
   (b) $p \preceq p$, $p \not\prec p$
   (c) For all p and q, $p \preceq q$ or $q \preceq p$
2. If $p \wedge q = \bot$ then
   (a) If $p \prec q$ then $p \prec (p \vee q) \prec q$
   (b) If $p \approx q$ then $p \approx (p \vee q) \approx q$
3. If $p \wedge q = \bot$ and $p \approx q$, then if $(p \vee r) \approx (q \vee r)$ for some r with $r \wedge p = r \wedge q = \bot$ and $p \not\approx r$, then $(p \vee r) \approx (q \vee r)$ for every such r.
4. If $p \prec R \prec q$ and $R = \bigvee_\alpha r_\alpha$ then for some β, $p \prec R_\beta \prec q$ where $R_\beta = \bigvee_{\alpha > \beta} r_\alpha$. Similarly if R is the infimum $\bigwedge_\alpha r_\alpha$.

Together these place a substantive constraint on the kinds of theory you can have about vagueness related attitudes. According to certain views it is hard to reduce preferences to the agents bouletic and doxastic attitudes. For example, these axioms make trouble for the project of reducing preferences to expected utilities in Field's framework. The first two reductions we mentioned in in §3.2.2 fail the linearity axiom, and the third reduction fails averaging.

3.3.3 Representing uncertainty

With a theory of preference in place we are in a good position to evaluate the status of vagueness related uncertainty. The strategy here is that in evaluating the judgments

concerning the preferences of an agent who is certain that p is vague, we will get some insight into the doxastic attitudes the agent has towards p itself.

Consider the following scenario. You have just won some unknown amount of money which you know to be between £N and £M. You also know that it is borderline whether you are now rich, given that you now own an additional N to M pounds. It seems as if things are better for you on the assumption that you are rich than they are for you on no assumptions. That is to say, you would prefer to find out that you are rich that to find out nothing at all: $\top \prec R$.[29] Similarly, things are worse for you on the assumption that you're not rich: $\neg R \prec \top$.

The reason for these preferences can be explained by appealing to the agents doxastic states. The first preference could be explained by appealing to the fact that conditional on the assumption that you're rich the hypotheses saying that you have a smaller amount of money, closer to £N, are not very likely whereas the hypotheses saying that you have a greater amount of money, nearer £M, are more likely on the assumption that you're rich. Symmetrical reasoning explains the second preference. So it seems that preferences can give us some insight into the nature of uncertainty, and in this case, seems to suggest that you should be uncertain whether you're rich *even though* you know that it's vague whether you're rich.

In fact, if we think of the preference relation as being coextensive with the relation of comparative conditional expected utilities – i.e. we think that $p \prec q$ iff $E(u \mid p) < E(u \mid q)$, where E is the conditional expectation operator, and u is a random variable for utility – then the above pattern of preferences can only hold if you are uncertain whether you're rich. In general if $\neg R \prec \top \prec R$, then you must be uncertain whether R: the fact that $\top \prec R$ guarantees you're not certain in R, and $\neg R \prec \top$ guarantees that you're not certain in $\neg R$. Note that since the claim that I'd prefer to be rich than not seems a very natural thing to say in this situation, we appear to have an argument against Field style views.

The above is only a condition on preferences which is sufficient for being uncertain. For necessary and sufficient conditions we can define uncertainty, in a preference theoretic sense, as follows.[30]

> An agent is $\prec$-certain that p if and only if $\forall q, r(q \prec r \leftrightarrow q \prec_p r)$ (and similarly for $\preceq$.)
>
> An agent is $\prec$-uncertain whether p if and only if she is neither $\prec$-certain that p nor $\prec$-certain that $\neg p$.

You are therefore certain that p in the preference theoretic sense if your unconditional preferences are the same as your preferences conditional on p – i.e. you tend to prefer *as if* p were the case, and you are uncertain in p when you are neither certain in p nor $\neg p$.

Intuitive judgements about preferences give us a new notion of uncertainty, namely, that of preferential uncertainty. Preferential uncertainty regarding p is a pattern of preferences which, if your preferences *were* represented by credences and utilities, would be represented with credences assigning intermediate values to p. Although both Field and Schiffer may urge that in cases where we are certain that p is vague we are not genuinely uncertain about p, they may still say, and may even be committed to the claim that we are preferentially uncertain in p. Such theories may respond to the problem of giving a sensible decision theory by separating the notion of genuine uncertainty from that of preferential uncertainty.

However, in many ways the notion of preferential uncertainty is more useful than the kinds of doxastic states theorists like Schiffer and Field propose, since it is this notion that plays the important role of informing the way we act in conjunction with our desires. Furthermore, it is unclear what words like 'credence', 'probability' and 'uncertainty' mean

[29] Of course, it is impossible to both find out that you're rich while also knowing that it's vague whether you're rich. But you can still evaluate how good things are for you on the hypothesis that you're rich.

[30] This definition plays an important role in some theories of knowledge. See Fantle & McGrath [31].

if not that thing which plays a certain functional role with respect to our desires, evidence and actions. On a functionalist understanding of words like 'belief' and 'credence' it is very hard to systematically separate genuine uncertainty from preferential uncertainty in the way that Field and Schiffer would have to do.

Earlier we established that given my preferences about being rich, I prefer as if I were uncertain about whether I am rich. Using judgements about preferences we obtained a fairly coarse grained description of my credential state. In fact we can in general recover quite a large amount of information about what an agents credences would have to look like from her preferences. This is seen by the following theorem due to Bolker:

Theorem 3.3.1. *(*Bolker*.) suppose that $\langle \prec, \preceq \rangle$ is a preference relation over a complete atomless Boolean $\mathbb{B}$ satisfying the Jeffrey-Bolker axioms. Then there exists a countably additive probability measure, Cr, on the propositions, and a utility function, u over the atomic propositions such that*

- *$p \prec q$ iff $E(u \mid p) < E(u \mid q)$ for $p, q \in \mathbb{B}$,*
- *$p \prec q$ only if $E(u \mid p) < E(u \mid q)$ for $p, q \in \mathbb{B}$, if the preference relation obeys the Bolker-Jeffrey axioms other than linearity.*

Here $E(x|p)$ represents the conditional expectation of the random variable x given p for the measure Cr. In the former case, if u is unbounded for any such representation, this representation is unique up to affine transformations of u.

This result is quite striking given the functionalist perspective on credences. What this result effectively tells us is that if a persons preferences conform to the Bolker-Jeffrey axioms then we can always associate with them a utility function and a credence function obeying the probability axioms in such a way that their preferences conform to comparisons of conditional expected utility. Assuming that credences just *are* whatever plays this specific functional role with respect to your preferences this makes for a powerful argument for probabilism.

3.3.4 Decision theory

So far we have said little about the relation between the theory of preferences and actual decision theory. There are a number of delicate issues surrounding the relation between a theory of rational preferences and a theory of rational decision.[31] In order to discuss decision theory I shall make use of a somewhat technical term: that of an agent having made it the case that p, which means, roughly, that p and that p came about as a result of the agents actions. One important issue that needs resolving is what it means for p to have come about as 'a result' of the agents actions; for example, to what extent must we invoke causal and intentional phenomena to explain this? A second important issue concerns which propositions can be taken as possible candidates for the agent to make the case; call these the 'action propositions'. Since the agent made it the case that p entails p, it follows that p had better be practically possible for the agent, but getting a clearer grip on the properties of action propositions is tricky. Finally, assuming these issues have been resolved, we need some way of connecting facts about preferences to what you should make the case. The simplest account simply states that if p is an action proposition and is preferred over every other action proposition then you should make it the case that p (and that if p is weakly preferred to every action proposition it is permissible to make it the case that p.) I shall not address these problems in a satisfactory way here, but having them in mind in what follows is important.

[31] See, for example, the literature surrounding Newcombe's problem [97].

Vagueness and Action

We have so far been discussing the properties of a certain binary propositional attitude, preference. It is very natural to want to know what kind of behavioural upshot our preferences have. In particularly one might want to know how one's preferences in vague propositions could have any distinctive behavioural consequences. On the one hand it would be bad if your vague preferences didn't contribute anything to your behavioural profile. It raises the question of whether these preferences are redundant and whether everything of importance could be captured by your preferences in the precise propositions. On the other hand it would be curious if your preferences in the vague did rationally contribute to your behaviour since propositions about your actions are typically determinately true or false – the puzzlement is hard to verbalise, but there seems to be something odd about the idea that your preferences over vague matters could manifest itself in precise physical behaviour not explainable in terms of your preferences in the precise.

What kind of preference between two vague propositions could manifest itself in a behavioural fact which could not equally be explained by one's preferences between precise propositions? In chapter 2 we discussed the possibility of two agents, Alice and Bob, having the same credences in every precise proposition but having different credences in the vague. If additionally the things they cared intrinsically about are they same, in the sense that their utility functions are identical, then their preferences would only differ on vague propositions. Would it follow that Alice and Bob were behaviourally identical? In this section I shall argue that this does not follow.

Here is where our preferred theory of actions helps to clarify things. On this view your actions are the propositions you have made the case – there is no restriction to the effect that the propositions you've made the case must be precise, however. Note that 'makes it the case that' is plausibly closed under logical implication so that you have made the case all of the consequences of the things you've made the case. I shall therefore stipulate that for a proposition to be among your action propositions it must be possible, in some restricted sense, both that: (1) you've made it the case that p and (2) that for any q, if you've made it the case that q then p entails q. That is, for p to be an action proposition it must be possible (in some sense) that it be the strongest proposition you've made the case. In general there will always be a strongest such proposition provided 'you make it the case that' is closed under conjunctions. Note that the set of action propositions is in general not required to be a partition, the tautologous proposition is not always an action proposition[32] and we may even countenance two distinct action propositions such that one entails the other.

When we are considering the behaviour of preference with respect to vague propositions there are even more issues to take into account concerning the relation between preferences and decisions. For example, we argued that in a situation where I know I have between N and M pounds and I know that it's vague whether someone who has any amount of money between N and M pounds is rich, things are better for me on the assumption that I'm rich than on the assumption that I'm not: $\neg R \prec R$. However, being rich may simply not be among my action propositions. I may be able to control whether I'm rich or not by by working harder or making good investments; but in such cases the strongest proposition I've made the case will typically be some precise proposition which entails that I'm rich, such as for example, the proposition that I have £100,000. In such cases the proposition that I'm rich is not the strongest proposition I made the case and is therefore not an action proposition.

It is hard to imagine what one would physically have to do to be such that the strongest thing that you've done is made yourself rich. A question therefore suggests itself: could a potentially vague proposition ever be an action proposition; i.e. could a potentially vague proposition, p, ever be such that you've made it the case that p and for every q if you've

[32]This just follows from the fact that it is some times not possible to do nothing.

made it the case that q then p entails q?

On the one hand it is very natural to think that we must permit vague propositions into the action set. In almost all decisions actions are macroscopic events, and states of affairs described in macroscopic terms are almost always vague. A prohibition against considering decisions in which the action propositions are vague propositions would be overly restrictive. Suppose I think I will miss my train; my options, at a first parse, are to run or to walk to the station. But clearly there are Sorites sequences starting with situations in which I'm walking and ending in situations in which I'm running showing that the proposition that I run is potentially vague. It might *seem* that whatever events in fact unfold are determinately true or false: my running is necessarily equivalent to some conjunction of precise microphysical facts, M. But I didn't *make* M the case, or least M isn't the strongest thing I made the case, even though I made myself run and necessarily I run iff M. This is because making the case is a partly intentional notion; to make p the case it presumably has to be some kind of causal consequence of something I intended to do, and 'intending' is hyperintensional (as is the macroscopic notion of causation, perhaps.)

However there is an argument, due to Cian Dorr, that the action set of any determinately rational agent consists only of propositions which are determinately true or determinately false if it's determinate that there are no ties among her preferences on actions. Let x be a determinately rational agent, A her action set and let $\preceq$ represent her preferences.

1. For any pair of propositions p and q: either it's determinate that $p \preceq q$ or it's determinate that $q \preceq p$.
2. For any proposition p either it's determinate that $p \in A$ or it's determinate that $p \notin A$.
3. For any $p \in A$, determinately: p if and only if $q \prec p$ for every $q \in A \setminus \{p\}$.
4. Therefore for any $p \in A$ either it's determinate that p or it's determinate that $\neg p$.

The first and second premise are debateable, but are probably true in many relevant cases; for example the first premise reflects the thought that it is a precise matter what credence and utility the agent assigns to a proposition. (3) follows from the claim that the agent is determinately rational and there are no ties in A. If the agent is determinately rational and there are no ties then she makes the best proposition in her action set true. Finally from the precision of the right hand side of (3) we may infer that for every $p \in A$, determinately: p iff [something determinate]. This ensures (4).

What this argument delivers is that A must consist of propositions that are in fact determinately true or false; it may still contain potentially vague propositions, as we have argued above. Note that it is not possible to necessitate this argument to conclude the propositions are necessarily determinately true or determinately false. For each agent, even if they are in fact rational and determinately so, it is usually contingent whether they are rational. But more importantly, it's contingent what an agents action set *is*, and it is contingent what she knows and therefore contingent what her preferences are. Let A be her actual action set. Then, even if x were necessarily rational, there could be worlds where the best proposition in her action set at that world was true, even though the best proposition in A is not true. Similarly, there could be worlds where her action set is the same, i.e. A, but her knowledge and preferences are different so that at this world the best proposition in A is true, even though it's not the best proposition according to her preferences at the actual world. Suppose that in order to catch the train at this world it'll be determinate that I run provided it's determinate that I'm rational. In worlds where I move towards the station in a way such that it's vague whether I run, either (i) I am not determinately rational at this world, or (ii) I know I'll catch the train even if I move in this way (i.e. my knowledge and preferences are different) or (iii) my moving this way is somehow beyond my control even though I am both rational and running is the best proposition in my action set.

To return to our original problem: how could Alice and Bob rationally behave differently given that all their disagreements are about the vague? We can in fact see that this difference can manifest itself in different ways. Alice and Bob may have different action sets; there may be propositions which Alice can make the case but Bob can't, for example. But even if their action sets are the same, it could be that the best proposition in Alice's set is different from the best proposition in Bob's set. This would imply that either Bob's or Alice's action set contains a vague proposition since they agree about the ranking of all the precise propositions. Thus, assuming that Alice and Bob are both rational, this will result in them making different propositions the case (even if they look like they're doing the same thing.)

Vagueness and Betting

Many philosophers, some of whom we have discussed already, have argued that probabilism is false due to issues surrounding vagueness related attitudes. In fact it is sometimes claimed that there is a decision theoretic argument against probabilism: seemingly rational agents do not accept bets on propositions they know to be borderline even if they are favourable according probabilistic calculations (Dietz [23] and Smith [124], among others, push various versions of this line of thought.) Since an argument against probabilism is, *ipso facto*, an argument against Jeffrey-Bolker decision theory, it is important that I address this issue.

A natural way to measure the degree to which someone believes a proposition is to look at how they behave. As a rule of thumb, if a persons credence in a proposition p is x, then they will usually be willing to accept a bet that costs $£x - \epsilon$ which pays out £1 if p and nothing otherwise. Here $\epsilon > 0$ can be made arbitrarily small. A Dutch book argument uses facts like this, and premises about what kind of betting behaviour is rational to argue that an agent must be probabilistically coherent. A Dutch book argument is like a behaviouristic version the representation theorem in section 3.3.3: rather than showing an agents credences are determined from her preferences, taken as a *sui generis* kind of propositional attitude, we show how a rational agents credences are determined by her behaviour. However preferences can hold between propositions which are not action propositions: in the example in which I knew it was vague whether I was rich the proposition that I'm rich was not among my action propositions (even though precise propositions entailing that I was rich probably were), so it is not clear that preferences and dispositions to act will match up. Indeed, this is what we shall show in this section.

While the Dutch book argument might be compelling in ordinary contexts, it is not so clear that the argument can carry over to contexts where the bets in question are bets on propositions you are certain are indeterminate. There seems to be something deeply wrong about participating in a bet concerning Harry's baldness when you know that it's vague whether he's bald.

The fact that an agent who is certain that p is vague will typically neither accept a bet on p or on $\neg p$, even if they both pay out £1 and both cost 10 pence to play, seems to suggest that an agents credences will fail to be probabilistically coherent – at least, if we accept the behaviouristic account of credences. Dietz [23] has used a similar argument to provide a Dutch book argument for Field's probability calculus.[33]

What are we to make of such bets? The first misunderstanding I wish to dispel is the thought that a bet on a potentially vague proposition is somehow a defective bet in the same way as, for instance, a bet on a question is defective. Let's say a bet is defective in this kind of sense if there's no well defined advice you could give someone facing such a bet. Since bets are rarely ever made on matters which are necessarily precise almost all

[33]However Dietz simply stipulates that you lose a bet if the proposition in question turns out to be indeterminate. This looks suspiciously like a bet on the proposition that it's determinate that p, not on the proposition that p. If what he has in fact shown is that one's credence in the determinacy of a disjunction of incompatible propositions is not the sum of one's credences in the determinacy of the disjuncts, then we should not be worried; this upshot is compatible with finite additivity.

bets will be potentially vague; but it is clear that we can give advice on a great number of bets. I think this point generalises to bets on propositions which are in fact indeterminate. Imagine ten people are assigned a number between 1 and 1,000,000,000, and they're given a bet that they'll be assigned a big number. Let us suppose one and only one of the ten people are assigned a number which is such that it's vague whether it's big. I think it should be clear that this persons bet is defective only if everyone's bet is defective: since they are all in exactly the same epistemic situation any advice that applies to one of the ten people applies to them all.

I think the only sense in which a bet in a vague proposition is defective is the same sense in which a bet on an unverifiable proposition is defective. For example, it would be a bad idea to measure someone's credences in propositions concerning events outside our lightcone by observing their betting behaviour because the winning conditions for such bets would always be unknown. No rational person would accept either kind of bet, vague or unverifiable, at odds matching their credences because no rational person could be certain that the bet's payoff will correlate with the truth of the proposition in question. In these examples you have reason to believe you're not in the kind of decision problem to which the decision theory outlined would apply. The examples *do not* show that the decision theory I've outlined fails to apply in the cases where it should.

There is, I think, a much simpler explanation as to why a person who knows p is vague would typically be disposed to reject bets on p with favourable odds and be disposed to reject bets on $\neg p$ with favourable odds. No agent would conditionally believe the biconditional: 'p if and only if I recieve the payoff' on the assumption that they accept the bet. Assume the agent has accepted; since it's always a determinate matter whether the agent receives a payoff and p is vague the biconditional is at best vague and at worse false. Since you can know this *a priori* you should not conditionally accept the biconditional.

In Dietz's set up you automatically lose a bet if the outcome turns out to be indeterminate. Note, however, that p may be indeterminate even though p, just as p may be contingent even though p. The the conditions under which you win one of Dietz's 'bets on p' are not the conditions under which p hold, but the conditions under which a completely different proposition, determinately p, hold.[34] According to Dietz you may lose a bet on p even when p.[35] Why should someone's betting behaviour with regard to such bets tell us anything about their credences in p?[36]

In short, the dispositions of a rational better do not reveal their credences in cases involving vagueness. This does not undermine the representation theorem we outlined in the previous section – the dispositions of a rational better do not reveal their preferences either.

The objection, recall, is that we bet as if we have non-probabilistic credences. In order to get the objection off the ground we need to make sure we're certain that we're in the decision problem alleged (e.g. we need to be certain the truth of p correlates with the payout.) Is it possible to modify the cases so that there is such a knowable correlation?

In suitably constructed cases I think it is possible to be certain that your payoff will be correlated with the truth of the proposition you are betting on, even if you're certain the proposition you are betting on is vague. Consider the following example. Over a period of

[34]Note that Dietz is using the word 'true' roughly as I am using the word 'determinate'. In a setting in which 'true' is not being used transparently we must distinguish carefully between the proposition that p and the proposition that p is true.

[35]Suppose that Alice accepts a bet from Dietz on p, and Bob accepts the same bet but on $\neg p$. p turns out to be borderline, so they both lose. It follows by classical theorems and modus ponens that either Alice lost bet on p even though p or Bob lost a bet on $\neg p$ even thought $\neg p$.

[36]It appears that Dietz is motivated by the thought that we should be able to verify the outcome of a bet. As a practical measure it is unclear whether only considering bets on the determinacy of a proposition makes matters better: just as it can be indeterminate whether p, it can be indeterminate whether p is determinate. Similarly, claims about events outside our lightcone are determinate but not verifiable. This aside, our betting behaviour towards the proposition that p is determinate does not license any conclusions about our credences in p other than, perhaps, weak constraints like $Cr(\Delta p) \leq Cr(p)$.

4 years Fred revises his will 17 times at regular intervals. At the beginning of the 4 years he is perfectly sane, and by the end of the 4 years he is clearly insane. At the time of Fred's death, the most recent valid will determines the transfer of money and property to any beneficiaries. Suppose that a will written by a mad man is not legally valid, and we may assume this is the only reason any of the wills might fail to be valid.

Now suppose at some point at which he is bordering on madness Fred offers you a bet over whether he is currently mad. The bet costs £50 to participate in: Fred says he'll guarantee that you will legally own £100 if he's mad and if not you'll get nothing. How does Fred guarantee that you'll legally own the £100 just in case he's mad? It is simple – he writes it into his will. We may assume that when he dies it is indeterminate whether the £100 belongs to you, or to whomever it is promised in the other versions of his will. However, due to the nature of the situation, it is determinately true that you will legally own the £100 if and only if he was mad. Thus you may be certain that the bet is not defective; you are certain that you'll own the money if and only if he's mad.

So in general it seems that we can, at a stretch, describe scenarios in which the payoff of a bet is penumbrally correlated with the propositions we are betting on. There is a further problem however. Why should anyone care about legally owning money if they can't spend it on the things that they really want? In order to get around this we could just stipulate that the agent owned about legally owning money and not about the concrete consequences that normally come along with legally owning money.

These cases are somewhat odd since they require the agent to care intrinsically about the truth value of a vague proposition (in this case to care intrinsically about owning money, even when it's vague whether you own the money) – something I will ultimately argue is irrational. However, bracketing that issue for a moment, the real question is: do these cases make trouble for probabilism? It seems to me that they do not. Once we have been careful to ensure that the agent really does believe she is in the decision problem described, and that she really cares intrinsically about the payoff, I no longer have the intuition that she should reject favourable bets on p and $\neg p$.

3.4 May we care intrinsically about the vague?

The final examples considered in the last section relied on an assumption, not particularly plausible, commonly made in Dutch book arguments: that people care intrinsically about legally owning money (and only care about this.) This assumption is innocuous only if the arguments would go through if money were replaced throughout with something the agent cared about. In order to make the Dutch book arguments work we needed to construct cases where it was vague whether you owned money. Is it possible to substitute money for something you can actually rationally care about, while keeping the feature that you know that the outcome you care about is penumbrally connected to the vague proposition you're betting on?

The issue at stake, therefore, is whether it is possible to rationally care intrinsically about the vague. In the last section we showed that *if* someone cared intrinsically about the vague we could construct something similar to the ordinary Dutch book arguments for probabilistic coherence. But there is still a question of whether it is possible, rationally, to care about vague things. I shall argue that ultimately whether it is rational to care about the vague, and whether the Dutch book arguments work, depends on your views about the nature of vagueness.

To demonstrate what this might amount to we might consider an example. Suppose Harry has the desire to be bald, and, for simplicity, let us suppose this is the only telic desire he has. He could have this desire in a number of ways – for example, he might have the desire derivatively in virtue of his having the desire to have no hairs at all. Let us suppose that this desire is not had derivatively, it is the strongest proposition about his head that he desires: Harry wants to be bald and any other state of hairiness that he wants is already

entailed by his being bald. There is therefore no specific number of hairs he wants; if he wants to have 98 hairs or less this is only because he thinks this will make him bald and not because he cares about having smaller number of hairs over larger numbers. He does not care whether he has 98, 99, 100, ..., hairs unless it would make the difference between his being bald and his being not bald – that is, he cares about his numerical hair number only insofar as it contributes to making him bald.

Compare this with another case. Alice cares intrinsically about whether there is intelligent life elsewhere in the universe right at this second. For simplicity, again, we may assume this is the only thing she cares about. Like Harry, she doesn't really care about other features of the universe. She is also certain that if there is intelligent life it's well outside the region spanned by her lightcone at any point in her life. Whether or not there is intelligent life out there is completely causally independent of her and is something should could never hope to verify. Nonetheless, she thinks that if there is intelligent life out there that would be a pretty cool thing.

I think there are two kinds of reactions you could potentially have to these two examples.

> Reaction 1. Harry's desire is just like Alice's desire. They both desire something which could never have a causal effect on either of them. The truth values of both the propositions they desire will never be known. The fact that Harry is bald (is not bald) is just like the fact that there is (is no) intelligent life in the universe. While they are both unreasonable desires to have, in some weak sense, they are perfectly coherent.
>
> Reaction 2. Harry's desire is not at all like Alice's. While their desires are both unknowable and causally independent of their immediate surroundings Harry's desire is simply incoherent. Harry cares about whether he is bald or not when he already knows exactly how many hairs he has. Based on this desire Harry acts as if there is a fact of the matter out there, whether he's bald, just as there's matter of fact about whether there's intelligent life in the galaxy; only Alice is entitled to act like this.

It is not unnatural to associate the first kind of reaction with an epistemicist who think that vagueness is *merely* a matter of ignorance, albeit a special kind of ignorance. For the epistemicist facts about baldness in borderline cases are not especially different in kind from facts about life in the far reaches of the galaxy. One consequence of this, it is natural to think, is that there is no distinctive reason why we could not care about whether a borderline case of baldness is bald or not.

The second kind of response is naturally associated with supervaluationists and other theorits who think that in borderline cases there simply is no fact out there to be known, and no fact worth caring about. Propositions about life in the galaxy are different since, while they are like borderline cases in the sense that they are unknowable, there is still a fact of the matter *out there* which one could reasonably care about. For this latter kind of theorist Alice's desire is perhaps a little eccentric, perhaps even irrational in a loose sense, but it is not incoherent. Bob's preferences, on the other hand, seems to be conceptually confused.

This observation is intended to help us with the problem we raised at the beginning of the chapter and in section 3.1. If epistemicists and other factualists think that vagueness is *just* a matter of uncertainty, then what is this extra thing non-factualists think vagueness involves? According to the present suggestion, if you know there is no fact of the matter whether p then you shouldn't care whether or not p. We shall spend much of this section sharpening this claim.

It is worth contrasting this with a difference between a factualist view and the particular non-factualist view I have been defending so far. Suppose Alice and Bob know that Harry has exactly N hairs, and that N is in the borderline region. I have argued that if Alice and Bob are conceptually coherent they will have the same credence as each other in the proposition that Harry is bald. Although a factualist will agree that Alice and Bob should

have an intermediate credence in this proposition, it seems that there could be reasonable disagreement between them about whether Harry is bald – they could assign different credences to this proposition. This would be just another respect in which the claim that Harry is bald is like the claim that there is intelligent non-human life in the Milky Way: in both cases two people could reasonably have the same evidence yet have different credences. It is not clear that every non-factualist view need agree with me that your rational credences in the precise fix your credences in the vague, but it is worth noting that, for the present view at least, there is something else I can say to distinguish myself from the factualist.

Let's consider the concrete consequences these two theses have. It would hardly be interesting if the thesis that one should only care about the precise had no tangible consequences. Fortunately, if we assume the supervenience thesis, the requirement that we not care about the vague it is not merely epiphenomenal, there are concrete differences between a factualist view and the view I have been defending: there are decision problems where people who differ in whether they care about the vague make different actions.

For example, the decision problem we considered in chapter 2 has this property. Let us recall the example:

> Alice and Bob are in an auction house and they are both bidding on a relatively inexpensive vase which is a shade of turquoise that is indeterminate between being blue and green.
>
> Now suppose we know the following things about Alice and Bob. Firstly, they have exactly the same telic desires and have the same credences in the precise propositions. We also know the following facts. (1) as it happens they are both very peculiar people and they care intrinsically about owning green things. (2) although they both know exactly what precise shade the vase is, Alice is more confident than Bob that the vase is green.

Note that this kind of scenario is simply impossible according to the view I have been defending. It requires that Alice and Bob care intrinsically about the vague, and that they disagree to some degree about whether the vase is green, even though they know the relevant precise facts such as its exact shade.

It is not hard to show, according to my view, that if Alice and Bob have the same telic desires and they both know all the relevant precise facts they will always act in the same way if they are rational. This is not so in the above scenario: it is clear that Alice should bid more than Bob for the vase. Since they care about the same things its value is the same for both Bob and Alice. However, since they know that it is vague whether the vase is green they are uncertain whether it is green and therefore the purchase is a gamble on its being green. However, according to Alices subjective credences the odds that it is green are better than they would be according to Bob's. Therefore Alice ought to outbid Bob.

On the proposed view the kind of scenario described above requires Alice and Bob to have deeply irrational preferences, and at least one of Alice and Bob to have incorrect credences. On the other hand, the scenario seems fine for the kind of factualist I have been describing. The factualist thinks that one can reasonably disagree about a proposition, even given all the relevant precise evidence, and that one can care intrinsically about the vague. So this constitutes a fairly substantive difference between the proposed view and factualism about vagueness.

A natural question to ask next is whether we can run an example like this without relying on thesis that vague beliefs must rationally supervene on your precise beliefs? This would be important for distinguishing factualism from forms of non-factualism which reject the supervenience thesis. In this case things depend on the issues we raised in section 3.3.4. If Alice and Bob have the same credences, and have a utility function which assigns the same value to every precise proposition[37] then whenever p and q are precise $p \prec^A q$ iff

[37]It is of course controversial that we can make interpersonal comparisons of utility like this, but all we

$p \prec^B q$. It follows that if the action propositions for Alice and Bob are the same and consist only in precise propositions then Alice and Bob are behaviourally equivalent if they're both rational, in the sense that they have the same dispositions concerning which propositions to make the case.

However, in section 3.3.4, we argued that even if the set of possible propositions an agent can make the case consists only of determinately true or determinately false propositions, it may still contain vague – i.e. potentially indeterminate – propositions. Since Alice and Bob's preferences can differ over the potentially indeterminate propositions it follows that Alice and Bob may make different propositions the case. The upshot of all of this is: even without the supervenience thesis, factualist and non-factualist may have concrete disagreements about what kinds of behaviour is rational.

Suppose we change the example: this time Alice and Bob are both presented with a sequence of vases. Suppose the vases can be ordered into a Sorites sequence starting with a clearly green vase and slowly changing colour until the vases are clearly blue. However the way they are arranged for Alice and Bob is randomised. Suppose also that they know the precise shades of each vase, but they are in a rush and don't have much time to spend thinking about each vase. Now Alice and Bob may in fact chose the same vase, yet they may still be behaviourally different. For example Alice picked out a green vase, so Alice made it the case that she obtained a green vase, while Bob only went by the shades so Bob made it the case that he obtained a vase within an interval of specific (precise) shades I. The proposition that Bob made the case may be necessarily equivalent to the proposition that Alice made the case, but they are distinct nonetheless, so they are behaviourally different concerning what they made the case.

3.4.1 The indifference principle

It's obvious that you can care about matters which are vague: for example you can rationally care about being rich, being happy, having lots of friend, and so on and so forth. The proposition that you're rich or that you're happy, and so on, are all vague propositions. This is all compatible with the claim that one cannot care *intrinsically* about the vague. However these examples do demonstrate that we need to be careful how to spell out the thesis, as its plausibility is very sensitive to the way it is stated.

In order to do this we shall assume that the algebra of propositions we are working with is atomic. In a supervaluationist framework the atoms would not correspond to possible worlds, but to pairs of possible worlds and precisifications. If we are to work in a neutral setting in which propositions are taken as primitive, however, a state is a proposition such that for any proposition (vague or precise) it entails that proposition or its negation, but doesn't entail everything (or alternatively: entails everything it's entailed by.) We can also partition the set of states into maximally strong precise propositions: propositions w such that for every precise proposition, p, w either entails p or p's negation. In general the strongest precise proposition entailed by a state (i.e. atom) i is a maximally strong precise proposition. In chapter 1 I called the atoms indices or impossible worlds. I shall retain the neutral term 'maximally specific' proposition for the atoms, and 'maximally specific precise proposition' for latter thing.

With these notions in place we can state the principle that correctly encodes the thought that one should not care intrinsically about the vague[38]:

IP. If the strongest precise proposition p entails is the strongest precise proposition q entails then I should be indifferent between p and q for maximally specific propositions p and q.

need to assume for this discussion is that Alice's utility is a positive affine transformation of Bob's over the space of precise propositions.

[38]It should be noted that, since there is vagueness concerning which propositions are precise, there can be cases where it's vague whether you may care about p.

Note that this principle does not entail that you couldn't rationally care about being rich. For simplicity let's assume you only care about your financial situation and let's also assume that whether you're rich supervenes on how much money you have. Given these assumptions IP *does* entail that you should be indifferent between states that agree on how much money you have. If you furthermore know how much money you have, you should be indifferent between all the different maximally specific ways things might, for all you know, be, even if they disagree about whether you're rich or not.

3.4.2 Objections to the indifference principle

Before we conclude this chapter let me consider a number of objections to the thesis.

Objection: According to some theories propositions about the future, e.g. whether there will be a sea battle tomorrow, are indeterminate. Yet evidently there is nothing irrational about caring whether there will be a sea battle tomorrow.

Response: I reject the theories which posit indeterminacy in propositions like these. If you care about what will happen tomorrow then there is a fact of the matter out there to care about. I suspect that such theorists simply mean that the future is open in an epistemic sense, a thesis with which IP is compatible.

Objection: According to some theorists, if you under go fission, i.e. your body is somehow divided in two, Lefty and Righty, then it's indeterminate whether you are Lefty or Righty after the fission. But presumably it is permissible to care about whether you'll be Lefty or Righty for example, if one is to be tortured and the other rewarded.

Response: Here I am uncertain, but inclined to agree that one can rationally care about what happens to you in fission cases. At any rate, any fission case in which you are rational to care about what happens to you is one in which you are not certain it is indeterminate which person you end up as. I am much more confident, however, that whether there are fission cases involving indeterminate survival or not, there are no cases where it is *both* known to you to be indeterminate who you survive as, and in which it is rational to care which person you survive as.[39] (See Williams [135] for further discussion of this kind of objection.)

Objection: Couldn't one care intrinsically about being happy.

Response: IP is consistent with one caring about being happy – what IP rules out is that you care about being in states which count you as happy over states in which count you as not happy, even if they agree about all the precise facts. As a simple model, imagine that there are a number of different factors which constitute your being happy. When these factors are present in certain combinations, you are happy, and when they are present in others you are not; some combinations leave it indeterminate whether you're happy or not. One way to think about IP is that you should care about what the combination of factors is like, and not whether they count as sufficient for 'happiness' or not.

Objection: Pretty much the only concepts that can be used to express precise propositions and properties are concepts of mathematics and fundamental physics. It is firstly grossly implausible that I be required only to care intrinsically about the propositions of fundamental physics and mathematics. Secondly, it follows that it was only until quite recently, with the discovery of modern physical concepts, that we were able to have thoughts about these precise propositions at all.

Response: While I agree that *sentences* of a public language rarely express precise propositions it does not follow that we do not bear any interesting relations to precise propositions in the sense described by my preferred theory of propositional attitudes. Furthermore it is not the case that the propositions of fundamental physics (i.e. propositions expressed in the vocabulary of fundamantal physics) are precise propositions: even if electronhood, for example, is in fact a basic property of fundamental physics, it is not a precise

[39] Note that according to some views *the very fact* that you can rationally care whether you survive as Lefty or Righty is partly constitutive of your surviving as one of them [98].

property in my sense, since there are epistemically possible words in which electronhood is not fundamental and it is vague whether something is an electron. For all I know the correct physics may be one in which fundamentally only fields exist, and electrons and other seemingly fundamental particles emerge out of these in a way that allows for vagueness concerning what an electron is. There are consistent worlds in which this is true even if they aren't nomically or even metaphysically possible (see the discussion of section 1.2.2.) The notion of a precise proposition can be implicitly introduced by the role it plays in a theory of rational propositional attitudes, a theory in which knowledge of physics is not a requirement of rationality.

It is possible that the objection rests upon an equivocation with the word 'property'. There are two conceptions of properties: an abundant conception according to which properties are abstract objects, serve as the semantic values of predicates, feature importantly as the objects of belief and other intentional attitudes, and a sparse conception according to which properties are physical things whose existence depends on fundamental features of the world and which are discovered by science. The property of being an electron, in the latter sense, may in fact *never* feature in the thoughts of an agent, even an agent who has all the concepts needed to state a final theory of fundamental physics.

Chapter 4

Vagueness At Every Order

It is an upshot of classical logic that if there are any small numbers at all then there is a last small number. It is compatible with this result that it is a vague matter which number that is. The boundary between the small and non-small isn't precise. There is a boundary, but it's *vague* where it lies, and it is the existence of precise boundaries, not vague ones, that we should be worried about.

This is, in a very schematic form, the classical response to the Sorites paradox.[1] To illustrate why precise boundaries (but not vague ones) are problematic consider the following example. There is something very bad about asserting that the total length of your childhood was 378432178928476829 nanoseconds. Vagueness prevents you from ever discovering this, and similar precise facts about the length of your childhood. If the boundary between your childhood and the rest of your life was not vague, however, there would have been no reason you couldn't have discovered the length of your childhood in nanoseconds, just as, perhaps, one could find out the number of nanoseconds in a year, and no reason to refrain from going about asserting it. Everyone, regardless of her preferred account of vagueness, must agree that the above assertion is bad and that this badness is due to its being at best vague whether your childhood lasted for this length of time.

This reasoning extends. Say that it's determinate that p just in case p and it's not vague whether p.[2] Is there a precise boundary between the determinate children, in other words, the non-borderline children, and everyone else? It seems we should not be any happier about assigning sharp numbers to the length of one's determinate childhood than to one's childhood. There is a completely analogous Sorites for 'determinate child' as there is for 'child'. To be sure, there is a last child and a last determinate child in any Sorites sequence, but it is always vague which person that last child or determinate child is. Similar comments apply to the further iterations: it's vague which the last determinately determinate child is in the sequence, and so on and so forth through the finite orders.

Indeed, there are some people who are determinately^n children for any amount of iterations, n, and some which are not. Surely it is vague where that boundary lies as well? In other words, there are some children such that it's neither vague nor higher order vague (vaguely vague or vaguely vaguely vague or ...), whether they're children, and others such that it is either vague or higher order vague whether they are children, and it's indeterminate where the boundary between the two lies. To see this note that:

> The period of my childhood during which it was neither vague nor higher order vague whether I was a child was 378432178928476829 nanoseconds in length. (4.1)

[1] Although note that you want these claims to assuage the initial intuition that there can't be a last small number, know that you let the epistemicist off the hook too.

[2] I am thus using the phrase 'it's determinate that p' in a very neutral way which permits even an epistemic interpretation.

sounds just as terrible, and would sound terrible no matter what number one used. However, if neither this sentence, nor any sentence like it, were vague, what possible reason could prevent us from finding out whether or not (4.1) was true? If we knew (4.1) it would be hard to explain the inappropriateness of asserting (4.1) (although see Dorr [24] for a possible explanation.)

It should be noted that the only resources one needs to establish the existence of higher order vagueness is (i) an operator expression, 'it's vague whether', giving one the means to express when someone is a child without it being vague whether they're a child (that is, when someone is a determinate child, in my nomenclature), (ii) the acknowledgement that a Sorites sequence for 'x is a child' is typically also a Sorites sequence for 'x is a determinate child' and (iii) that any Sorites susceptible predicate has vague instances. Since all these considerations can be made independently of one's preferred analysis of vagueness this is a problem for everybody – epistemicism, for instance, offers no respite here.[3]

Some philosophers deny the phenomenon of even second order vagueness (see, for instance, [148] and [111].) According to these theorists, the duration of your determinate childhood is a precise length. However these philosophers do not typically offer concrete hypotheses about the exact number of seconds at which it begins to be vague whether you're a child (and at which you stop being a determinate child.) If this is due to some kind of inability on their part, some explanation of this inability is required. The explanation cannot be that the boundary is vague, because by assumption the distinction between being determinately bald and being vaguely bald is a precise one. One might reasonably wonder what this new kind of obstacle to knowledge is if it is not vagueness. Equally puzzling issues arise for those who think higher order vagueness cuts out at some finite level larger than 2 (see Burgess [17].)

The subject of this chapter concerns a number of arguments that purport to show that for any predicate, F, there must be a precise boundary between the things which are determinately F at every order (henceforth "determinately* F") and the rest. That being vaguely F at some order or other and not being vague at any are precise distinctions. If this argument succeeds we should expect to be seeing exact numbers associated with vague predicates all over the place. Indeed numbers that are in principle discoverable; thus one should not be surprised to hear things like 'my determinate* childhood lasted exactly 378432178928476829 nanoseconds' or 'I became determinately* bald after I lost my $1451st$ hair' and so on. It is tempting to sweep the infinitary version of the Sorites paradox under the carpet - to say that predicates of the form 'determinately*-F' are indeed precise but they're so esoteric we shouldn't worry about them. I think that recognising that this response involves the possibility of finding out propositions like (4.1) is a cost many would not be willing to pay.

Let me introduce some notation. I shall write ∇p to mean it's vague whether p, Δp to mean that p and it's not vague whether p, and $\Delta^* p$ to mean p and it's neither vague nor higher order vague whether p, which is to say p, it's not vague whether p, it's not vague whether it's vague whether p, and so on. The problems considered can be generated without iterating into the transfinite ordinals, so by 'higher order vague' I shall just mean nth-order vague for some finite order n. Accordingly these operators can be connected in terms of an infinite conjunction as follows: $\Delta^* p \equiv \bigwedge_{n\in\omega} \Delta^n p$.

Classical logic shall be assumed throughout the chapter. The claim that it is never vague whether something is determinate* is formally represented by the schema

$$\Delta\Delta^* p \vee \Delta\neg\Delta^* p \tag{4.2}$$

which can be split up into two claims:

[3]Eklund [29], for example, claims that his particular analysis of vagueness fairs better with respect to the problem of higher order vagueness. However, since he can make sense of the property of being bald without being vaguely bald, and he believes that Sorites susceptability involves vagueness it is hard to see how this claim stands up to this version of the problem.

$(+\Delta)$ $\Delta^* p \rightarrow \Delta\Delta^* p$.

$(-\Delta)$ $\neg\Delta^* p \rightarrow \Delta\neg\Delta^* p$.

The former principle is introduced and argued for in Williamson [138]. Whilst some have rejected it, notably Field [40], it is not our intention to do so here. In fact an argument for $(+\Delta)$ is given in §4.1.2. The latter principle has received less attention, although it is crucial for the problem. One way to prove $(-\Delta)$ is to argue for the Brouwerian principle for Δ

(B) $p \rightarrow \Delta\neg\Delta\neg p$

Although B is the most obvious candidate (see [142], [26] footnote 11, [32]) $(-\Delta)$ is also entailed by a class of weaker principles

(B^n) $p \rightarrow \Delta(q \rightarrow \phi_n)$

where $\phi_1 := \neg\Delta\neg p$; $\phi_{n+1} := \neg\Delta\neg(q \wedge \phi_n)$.[4] I shall discuss the motivations behind these principles in §2. To this list I would also like to add a principle, which I dub B^*:[5]

(B^*) $\Delta(p \rightarrow \Delta p) \rightarrow (\neg p \rightarrow \Delta\neg p)$

In this chapter I shall explain how one can resist the conclusion that there is a precise boundary between the determinately* F's by rejecting these principles. I generalize some recent considerations (Mahtani [87], Dorr [25]) that show that B fails to show that B^n can fail for any n. I then consider B^* and provide tentative reasons against and in favour of it. Arguments by Wright, Graff and Zardini that do not appear to rest on B are also considered. Finally I discuss some technical issues concerning the logic of vagueness.

The structure of the chapter is as follows. In §4.1 I outline the problem of higher order vagueness, the principles which generate it, and argue that it is B and its weakenings that are responsible. In §4.2 I outline a picture of the structure of higher order vagueness in which B fails, consider the relation between vagueness, assertability and knowledge and address some other paradoxes which appear not to rely on B. Some technical questions to do with the proposed logic of vagueness are considered in the appendix.

4.1 The problem of higher order vagueness

Many authors have thought that precise cut-off points re-emerge once considerations involving higher order vagueness are taken into account. The following passage by Mark Sainsbury is often cited in favour of this conclusion:

> Suppose we have a finished account of a [vague] predicate, associating it with some possibly infinite number of boundaries, and some possibly infinite number of sets. Given the aims of the description, we must be able to organize the sets in the following threefold way: one of them is the set supposedly corresponding to the things of which the predicate is absolutely definitely and unimpugnably true, the things to which the predicate's application is untainted by the shadow of vagueness; one of them is the set supposedly corresponding to the things of which the predicate is absolutely definitely and unimpugnably false, the things to

[4] The weakenings of B sometimes considered are the principles $\mathsf{B}^{n\prime}$: $p \rightarrow \Delta\neg\Delta^n\neg p$. These are slightly weaker (see footnote 5 for the frame conditions) but we shall see that the strengthening is more motivated in this context.

[5] In terms of Kripke frames, B^n corresponds to the condition that if x can see y, then there is a path back to x from y with at most n steps each step of which x can see, whereas B^* entails that x can see a finite path back but places no bounds on the number of steps it might take. $\mathsf{B}^{n\prime}$ is weaker that B^n stating only that there is a path with at most n-steps back, but that x needn't see any of these steps.

> which the predicate's non-application is untainted by the shadow of vagueness; the union of the remaining sets would supposedly correspond to one or another kind of borderline case. So the old problem re-emerges: no sharp cut-off to the shadow of vagueness is marked in our linguistic practice, so to attribute it to the predicate is to misdescribe it. [116]

Raffman, for example, describes this reasoning as 'decisive.' However, as we have seen, its conclusion is paradoxical. If the distinction between the people who are 'absolutely definitely and unimpugnably' bald is a precise one, we ought to be able to find out and say where it applies much like we can in principle find out if someone has less than 1000 hairs. This much is characteristic of precise distinctions.

Much turns on whether you accept classical logic or a non-classical logic. Sainsbury's conclusion that a given object either falls under a predicates realm of application, absolutely definitely and unimpugnably, or it fails to do so in some way, is equivalent to an instance of the principle of excluded middle. Perhaps there is some reason why this particular instance of excluded middle must be true, but Sainsbury's argument has done nothing to establish that. For the classical logician, however, the conclusion has no bite; what distinguishes a precise from a vague predicate is not whether it obeys the principle of excluded middle. The question is: could it be vague whether a sentence is absolutely definitely and unimpugnably true, without a shadow of vagueness?

Let us grant, for the time being, this talk of claims having 'shadow's of vagueness.' The argument could continue. Suppose it could be vague whether a sentence was absolutely definitely and unimpugnably true – true without a shadow of vagueness. The fact that it is vague whether or not the sentence has a shadow of vagueness, presumably counts as its having a shadow of vagueness. Thus it isn't absolutely definitely and unimpugnably true without a shadow of vagueness. If it's vague whether a sentence is true without a shadow of vagueness, it's not true without a shadow of vagueness.

This conclusion, however, is completely compatible with its being vague, in some cases, whether a sentence is definitely and unimpugnably true untainted by the shadow of vagueness. For according to classical treatments of vagueness it is consistent to say that something is a vague instance of a property without falling under that property. For example, suppose it is vague whether Harry is bald, and (consequently) that it's vague whether or not he's not bald. By excluded middle, either Harry is bald or he isn't. Therefore either Harry is bald and it's vague whether or not he's not bald, or Harry is not bald and it's vague whether or not he is bald (by the classical inference $p, q \vdash ((p \wedge r) \vee (q \wedge \neg r))$.) In either case we have a property, F, such that Harry vaguely instantiates it, but doesn't in fact instantiate it (F is 'not bald' in the first case and 'bald' in the latter.) Of course, there are no determinate examples of things which are vaguely F without being F, so there are no determinate examples of things such that it is vague whether they have a shadow of vagueness. If it's vague whether a claim has a shadow of vagueness, it's also second order vague (and by analogous reasoning, it's vague at all orders.) But it is also compatible with a classical treatment of vagueness that it can be determinate that there are G's without there being any determinate G's; it is no objection to this approach that we cannot find any determinate examples of the phenomenon we are interested in. To conclude that every predicate is precise at higher orders requires further consideration.

4.1.1 The problem of higher order vagueness

There is much in Sainsbury's argument which is left unexplained. What do the adjectives 'absolutely', 'definitely' and 'unimpugnably' add? What does he mean by truth without a 'shadow of vagueness.' In this section I shall interpret having a 'shadow of vagueness' simply as having vagueness at some order, and I shall use this to develop a more specific version of this argument.

Imagine that we are talking about the natural numbers less than 100 and we want to know which ones are small. Obviously, there's no sharp boundary between small and non-small numbers. So there will be the numbers which are definitely small, the numbers which are definitely not small, and the borderline cases in between. There's also no sharp boundary between the definitely small numbers and the rest either: there are numbers for which it's vague whether they're definitely small or borderline small. To put it another way, there are numbers which are definitely definitely small, and those which are definitely not definitely small, but there's a range of borderline cases between the two in this case as well. Similar points apply to the boundary between the definitely definitely small numbers and the numbers which are not definitely definitely small, and so on and so forth.

One can see that the set of small numbers, the definitely small numbers, the definitely definitely small numbers, and so on, gradually shrinks as the number of 'definitely's' increases; after all, being definitely F is generally a more stringent condition than being F. But this set can't shrink forever! The first set in this sequence clearly starts off with less than 10,000 members, so in this most generous case it can shrink at most 10,000 times before it becomes empty. For some N being definitely^N small is the same as being definitely^m small for any $m \geq N$.

Does this result mean there is some kind of sharp boundary between the determinately^N small numbers and the rest? Fortunately this doesn't follow from what we have said so far. We may, and indeed must, accept that there is a set of numbers which are determinately^n small for every n, and a largest such set, but we may also maintain that *it's vague which set that is*. The starting point of this shrinking process - the set of small numbers - was vague, so there is no reason to think that it won't be vague which set you end up with after the shrinking process is complete. In keeping with the denial of sharp boundaries, we may still hold that it's vague which number is the last determinately^N small number. The most we can conclude is that *if* a number is determinately^N small it is determinately^k small at all orders k.

In the following sections I shall be considering various proposals that supplement this argument to show there must be sharp boundaries. First let us make this a little bit more formal. The toy argument above made an essential appeal to the fact that the Sorites sequence considered was discrete. Not all Sorites sequences are discrete, however, for example, smallness over the rational numbers, or redness over a spectrum. A completely general argument can be found in Williamson [138] which shows that if something is definitely^n F for any number of iterations n, then it definitely is definitely^n F for every n. We may define this strong notion of definiteness using infinitary conjunction:

$$\Delta^* \phi := \bigwedge_{n<\omega} \Delta^n \phi$$

In order to prove the result Williamson assumes the principle that conjunctions distribute over determinacy

(D) $\bigwedge_{i<\omega} \Delta\phi_i \rightarrow \Delta \bigwedge_{i<\omega} \phi_i$.

The argument proceeds as follows. We firstly note the logical truth: $\vdash \bigwedge_{n<\omega} \Delta^n \phi \rightarrow \bigwedge_{n<\omega} \Delta^{n+1} \phi$, which is just an instance of conjunction elimination.[6] From (D) we can immediately infer $\vdash \bigwedge_{n<\omega} \Delta^n \phi \rightarrow \Delta \bigwedge_{n<\omega} \Delta^n \phi$, i.e.

$(+\Delta)$ $\vdash \Delta^* \phi \rightarrow \Delta\Delta^* \phi$.

[6] This may be proved from the principles C1-C3 below

C1. $\bigwedge_{i<\omega} \phi_i \rightarrow \phi_n$ for each $n < \omega$.

C2. $\bigwedge_{i<\omega} (\phi \rightarrow \psi_i) \rightarrow (\phi \rightarrow \bigwedge_{i<\omega} \psi_i)$.

C3. If $\vdash \phi_i$ for each $i < \omega$, $\vdash \bigwedge_{i<\omega} \phi$.

$\vdash \bigwedge_{n<\omega} \Delta^n \phi \rightarrow \Delta^i \phi$ for each $0 < i < \omega$ by C1. Thus $\vdash \bigwedge_{i<\omega} (\bigwedge_{n<\omega} \Delta^n \phi \rightarrow \Delta^{i+1} \phi)$ by C3. So finally $\vdash \bigwedge_{n<\omega} \Delta^n \phi \rightarrow \bigwedge_{i<\omega} \Delta^{i+1} \phi$ by C2.

It should also be noted that none of these principles are characteristically classical, so this result extends to a number of non-classical logics.

4.1.2 Is a conjunction of determinate truths determinate?

The most natural place to block Williamson's argument is to deny (D). This is Field's strategy in [40]. But is this denial at all plausible? I claim it isn't if we accept the following principle concerning vagueness:

> If each constituent of a sentence is precise then the sentence itself is precise. (4.3)

Suppose for reductio that $\bigwedge_{i\leq\omega}\Delta\phi_i$ but that $\neg\Delta\bigwedge_{i\leq\omega}\phi_i$. Since each ϕ_i and infinitary conjunction are precise, it follows that $\bigwedge_{i\leq\omega}\phi_i$ is precise by (4.3), i.e. $\Delta\bigwedge_{i\leq\omega}\phi_i$ or $\Delta\neg\bigwedge_{i\leq\omega}\phi_i$. By assumption it's not true that $\Delta\bigwedge_{i\leq\omega}\phi_i$ so $\Delta\neg\bigwedge_{i\leq\omega}\phi_i$ must hold. By factivity we have ϕ_i for each i and $\neg\bigwedge_{i\leq\omega}\phi_i$ - this is a contradiction by C3.

It is interesting to note that if you accept the principle B, you can actually *prove* that a conjunction of determinate truths is determinate from the logic of conjunction alone (see appendix 3.2.) The thesis of this chapter, however, is that once you have rejected B, you already have a satisfying response to the paradoxes of higher order vagueness available. It is to this principle we now turn.

4.1.3 Sharp boundaries from B

To show that 'determinately* small' is a sharp predicate we must show that it can never be vague whether something is determinately* small. We have so far shown that if a number is determinately* small then it's not vague whether it's determinately* small. If we could show that if a number is not determinately* small then it's not vague whether it's determinately* small, we would be able to get our conclusion from an instance of excluded middle and reasoning by cases: every number is either determinately* small or it isn't, and in either case it is not vague whether it's determinately* small. The arguments I shall consider work in favour of the completely general principle I have called $(-\Delta)$: $\neg\Delta^*p\rightarrow\Delta\neg\Delta^*p$. In conjunction with $(+\Delta)$: $\Delta^*p\rightarrow\Delta\Delta^*p$ we then have the problematic principle $\Delta\Delta^*p\vee\Delta\neg\Delta^*p$ stating that there is never vagueness concerning what is determinate*.

A simple way to close this gap would be to introduce the principle B:

$$\mathsf{B}\text{: } \neg p\rightarrow\Delta\neg\Delta p \tag{4.4}$$

Here is how you prove $(-\Delta)$ from B: we have $\neg\Delta^*p\rightarrow\Delta\neg\Delta\Delta^*p$ by B. By contraposing $(+\Delta)$ we have $\neg\Delta\Delta^*p\rightarrow\neg\Delta^*p$, so we also have $\Delta\neg\Delta\Delta^*p\rightarrow\Delta\neg\Delta^*p$ by necessitation and an application of K. So by transitivity of the conditional we have $\neg\Delta^*p\rightarrow\Delta\neg\Delta^*p$.[7] Indeed, it is exactly this principle which is assumed in Williamson's discussion of these issues in [142].

The axiom B is motivated by Williamson's fixed margin models described in [138]. Williamson describes the class of Kripke frames $\mathcal{C}$ which contains frames, $\langle W,R\rangle$, for which there is some metric over W, $d(\cdot,\cdot)$, and $\alpha\in\mathbb{R}$ such that Rxy iff $d(x,y)\leq\alpha$. The motivation for this semantics is roughly the same whether we are epistemicist or some kind of supervaluationist. We should think of the members of W as precise interpretations of the language with the metric representing the degree of similarity between two interpretations. For example, suppose two interpretations, x and y, agree on how to interpret every

[7] It is interesting to note that B is not so central for the non-classical theorist. The problematic principle is the rule: if $\vdash\phi\rightarrow\Delta\phi$ then $\vdash\phi\vee\neg\phi$. For example both Łukasiewicz and Field's recent logic have this principle. Once one has this principle and $(+\Delta)$ all the problematic classical theorems concerning Δ^* statements are provable. Things would be different, however, if we adopted the weaker rule: if $\vdash\phi\rightarrow\Delta\phi$ and $\vdash\neg\phi\rightarrow\Delta\neg\phi$ then $\vdash\phi\vee\neg\phi$.

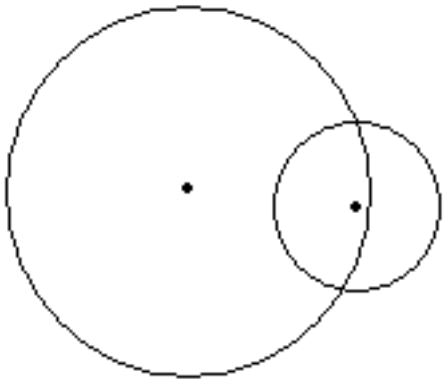


Figure 4.1: Mahtani's counterexample to B

expression, except x interprets 'small for a number less than 100' as the numbers less than 14 and y as the numbers less than 13. These two interpretations should be considered quite close by the intended metric, which is to say that $d(x, y)$ will be a relatively small number. A formula, ϕ, is determinately true according to an interpretion, x, if ϕ is true at all the interpretations similar enough to x. 'Similar enough' means the measure of their differences does not exceed α. At the interpretation x described above, '2 is small for a number less than 100' is determinately true because interpretations that disagree with x on this sentence are quite far away.[8]

It is clear that the accessibility relation of each such frame is symmetric: if the distance between x and y is less than α then, the distance between y and x is less than α. It is a standard fact that B is validated in all and only symmetric frames, so for any frame $\langle W, R\rangle$ in $\mathcal{C}$ the axiom B is valid.

4.1.4 Mahtani on failures of B

In [87] Mahtani argues that since the term 'determinately' is itself vague, its interpretation ought to vary from point to point and that Williamson's models fail to capture this fact. The interpretation of 'determinately' technically does vary from interpretation to interpretation on Williamson's semantics - being determinate at x depends essentially on which interpretations are closest to x. However they do not vary in all the salient respects. In particular, all interpretations agree about how close an interpretation has to be to be 'close enough' in the relevant sense; α is a fixed quantity throughout the model. In Mahtani's terminology the 'accessibility range' does not vary, when it should.

If each point, x, has its own accessibility range, $f(x)$, symmetry is no longer guaranteed. The distance between x and y may be less than $f(x)$ but not less than $f(y)$ (see figure 1.)

4.1.5 Sharp boundaries from other principles

Our initial concern was whether it must always be a precise matter whether something is determinate*. We noted that B was one way, but not the only way, to close the gap between Williamson's argument for $(+\Delta)$ and this claim.

Unfortunately there is an infinite chain of weaker principles, all stated in the finitary

[8]The closest interpretation to x which disagrees interprets 'is small for a number less than 100' as the numbers less than 2, whereas x interprets that as the numbers less than 14. In this context this constitutes a fairly big difference.

language, that also prove that 'determinate*' is precise.

$$\mathsf{B}^n\text{: } p \to \Delta(q \to \phi_n) \tag{4.5}$$

where $\phi_1 := \neg\Delta\neg p$; $\phi_{n+1} := \neg\Delta\neg(q \wedge \phi_n)$. And the principle

$$\mathsf{B}^*\text{: } \Delta(p \to \Delta p) \to (\neg p \to \Delta\neg p) \tag{4.6}$$

For example, adding B^* to the infinitary language allows us to prove the problematic $\neg\Delta^*\phi \to \Delta\neg\Delta^*\phi$. By applying necessitation to $(+\Delta)$, which we are assuming at this point, we have: $\vdash \Delta(\Delta^*\phi \to \Delta\Delta^*\phi)$. However, $\Delta(\Delta^*\phi \to \Delta\Delta^*\phi) \to (\neg\Delta^*\phi \to \Delta\neg\Delta^*\phi)$, is an instance of B^* in the infinitary language, so we can imediately infer $(-\Delta)$, i.e. $\neg\Delta^*\phi \to \Delta\neg\Delta^*\phi$, by modus ponens, as required.

In terms of frame conditions, B^n characterises the the property that if Rxy then there are n points, $z_1, \ldots, z_n$ such that Ryz_n, Rz_nz_{n-1}, ..., Rz_2z_1 and Rxz_i for each i - i.e. if x sees y then you can get back from y to x in n steps which x can see. B^* characterises the the property that if Rxy then for *some* n, there are $z_1, \ldots, z_n$ such that Ryz_n, Rz_nz_{n-1}, ..., Rz_2z_1 and Rxz_i for each i - i.e. if x sees y then you can get back from y to x in finitely many steps which x can see. Call this latter property the 'backtrack' property. In the lattice of modal logics, KTB^* is the infimum of $\{\mathsf{KTB}^n \mid n \in \omega\}$.

As stated, any one of these axioms is sufficient to close the gap between $(+\Delta)$ and the existence of sharp higher order cut-off points. I shall show that each frame validating KTB^n or KTB^* also validates $\neg\Delta^*p \to \Delta\neg\Delta^*p$ and hence $\Delta\Delta^*p \vee \Delta\neg\Delta^*p$. Suppose the frame $\mathcal{F} := \langle W, R\rangle$ validates KTB^n or KTB^*. For any model based on $\mathcal{F}$, if $\neg\Delta^*p$ is true at x and Rxy then (a) for some n you can get from x to a $\neg p$ world in n steps. (b) for some m you can get back to x from y in m steps. Thus you may get from y to a $\neg p$ world in $n+m$ steps, so $y \nVdash \Delta^{n+m}p$ and thus $y \Vdash \neg\Delta^*p$. But y was an arbitrary point accessible from x, thus $x \Vdash \Delta\neg\Delta^*p$. So $x \Vdash \neg\Delta^*p \to \Delta\neg\Delta^*p$ for every x.

It should be clear by now that to properly demonstrate that we can consistently deny higher order precision we will need a more principled way of generating counterexamples. This is what I shall attempt to do in the next section.

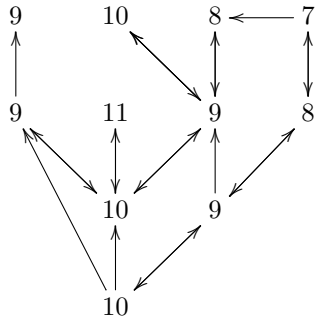
4.2 A B-free solution

To deny sharp cut-off points at all levels we must reject the principles B^n and B^*. The rest of this chapter is an evaluation of the prospects of this proposal. The general strategy is to find models in which the claims 'there are borderline cases of being determinately* (determinately^n) small' ($\exists x\nabla\Delta^*Sx$, $\exists x\nabla\Delta^nSx$) come out true. These models serve the purpose of showing that no contradiction can be derived from these assertions and plausible background assumptions. I attempt also to construct more realistic models, taking Williamson's fixed margin models as a starting point, to show that these principles are compatible with more realistic structural assumptions.

Below is a model which demonstrates that it is at least possible to deny sharp cut-off points at every level. Each node represents an interpretation of the language, with the number at each node representing the greatest number which satisfies 'small for a number less than 100' according to that interpretation. The truth value of each atomic statement of the form 'n is small' at a node is thus determined by whether number n is less than or equal to the number at that node. The truth values of extensional combinations of formulae are calculated as usual relative to a node, and a claim of the form '$\Delta\phi$' is evaluated true at a node x iff ϕ is true at every node accessible via an arrow from x.

No point can see a point which differs from it by more than one. The converse fails, however, since interpretations may differ radically in the interpretation of other expressions and might thus be inaccessible to one another. It is tacitly assumed that every point sees

itself.



Remember that 'the last small number' according to a point is the number written beside it, so 'the last determinately* small number' at a point is the smallest number you can get to from that point by following the arrows. Note that the bottom node can see a node where the last determinately* small number is 7: follow the arrow to the right. But it can also see two nodes where the last determinately* small number is 8: follow the arrow up or left. Thus it is vague, at this point, whether the last determinately* small number is 7 or 8. Indeed, it's vague whether it's determinate* that 8 is small.

Of course we tried to make this model look realistic by having several points (we could have gotten away with two) and making sure that adjacent points didn't disagree substantially over the interpretation of 'small for a number less than 100'. However, it would be nice to have a general class of models that includes such models as a special case, but are also constrained by facts about vagueness in the same way Williamson's semantics was. In fact we can modify Williamson's fixed margin models in just the way Mahtani suggests to allow for variation in accessibility range. We must also make sure that close points don't interpret 'determinately' drastically differently, i.e. we must make sure close points have similar accessibility ranges. This motivates the following definition.

Definition 4.2.0.1. *A v-frame is a triple $\langle W, d(\cdot,\cdot), f(\cdot)\rangle$ where $\langle W, d\rangle$ is a metric space, and $f : W \to \mathbb{R}$ obeys the following:*

(A) $\forall w, v \in W, |f(w) - f(v)| \leq d(w, v)$

A formula of propositional modal logic is valid on a v-frame $\langle W, d, f\rangle$ iff it is valid on the Kripke frame $\langle W, R\rangle$ where Rxy iff $d(x, y) \leq f(x)$. We shall talk about a v-frame and its associated Kripke frame interchangeably from now on.

The elements of W are to be thought of as interpretations or precisifications of the language, with the metric d representing how close they are to one another. A formula is determinately true at an interpretation if it is true at all nearby interpretations. What counts as nearby the interpretation w is determined by $f(w)$: v is nearby w when the distance between them according to d is less than $f(w)$. Note that what counts as 'nearby' depends on the precisification - the constraint (A) says, roughly, that the closer two interpretations are, the less they can differ over their interpretation of 'nearby'.

A useful fact is that there is a natural way to assign a metric over a (generated) Kripke frame. Simply assign a length to each arrow and define the distance between x and y to be the length, ignoring the direction of the arrows, of the shortest path between x and y. With a bit of fiddling one can show that the model above is generated by a v-frame in this way. The fact that one can refute B^n and B^* over v-frames follows also from the more general fact that the logic of v-frames is KT (see the appendix.)

4.2.1 Realistic frames

Let us consider a toy propositional language whose only atomic sentences are English sentences of the form: 'a is red', where a ranges over names for colours in a fixed colour

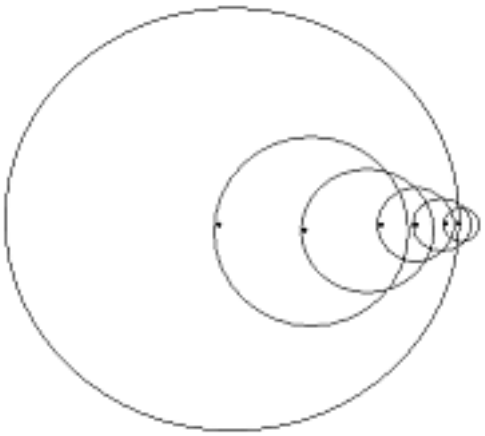


Figure 4.2: The validity of B^*: x sees a far out point on the edge, yet there is a finite path back to x.

spectrum. It is natural to suppose the interpretations of this language are completely specified by the cutoff point for 'red' along this spectrum of colours. Each colour can be represented by a real number, and the distance between two interpretations can be modelled as the difference between the two numbers representing the cutoff points for those interpretations. Thus the metric of our v-frame is the standard notion of distance on $\mathbb{R}$.

This suggests a very natural class of v-frames for modelling vague languages: those based on Euclidean space, $\mathbb{R}^n$, where the dimension n is the number of vague predicates in the language. The good news about these v-frames is that each of the problematic principles B^n are refutable. To refute B^n we consider the standard metric over $\mathbb{R}$. Let $\epsilon = \frac{1}{2^{n+1}}$. Stipulate that $f(0) = 1$, that $f(x) = f(-x) = \epsilon$ for $x \in \mathbb{R}\backslash[-1+\epsilon, 1-\epsilon]$ and $f(x) = 1 - x$ for $x \in (0, 1-\epsilon]$ and $x - 1$ for $x \in [\epsilon - 1, 0)$. It is easy to check this satisfies condition (A) and is a v-frame. Now 0 can see 1, yet the shortest path back from 1 takes $n + 1$ steps, thus B^n does not hold. Note that in this model there are no points which can only see themselves, i.e. no points, x, such that $f(x) = 0$.

The counterexamples traded on the idea that for any n one can find a v-frame based on an ϵ small enough to ensure that the longest path from 1 to 0 is longer than n. There is, however, no single model which refutes all the B^n simultaneously. Indeed it is not hard to show that v-frames based on $\mathbb{R}^n$ where no points have a 0 accessibility range has the backtrack property: if x sees y, then there is a finite path $z_0, \ldots, z_n$ such that y sees z_0, z_0 sees z_1, $\ldots$, z_n sees x and x sees each z_i. Thus it follows that B^* holds in these frames (see figure 2.) Intuitively, the closer a point is to x, the closer in diameter its accessibility range must be to x's thus one always can find a path leading back to x:

For our purposes this result would be devastating. It would entail, for example, that being determinately* red had completely precise boundaries, and this brings along with it all the problematic consequences already mentioned.

An obvious place to resist the result is to deny that the accessibility range of any point must be non-zero. Relaxing this constraint is independently motivated. Suppose we are working with the toy language described above, and the ordering of the relevant colour spectrum is not only dense but complete in the sense of containing a limit for any converging sequence of points (thus, for example, its structure is like that of $\mathbb{R}$, but not of $\mathbb{Q}$.)[9] If one made the assumption that $f(x) > 0$ for any interpretation x, it follows

[9]If one were to object that some fact about colours prevents the existence of such a spectrum, we could modify the example to be about the vague predicate 'small' as applied to real numbers.

that *no* colours in the spectrum are determinately* red. It is sufficient to show that any two interpretations, represented as real numbers, x and y, can be connected by a path. Without loss of generality we may assume that $x < y$. Let $(a_n)_{n\in\omega}$ denote the sequence $x, x + f(x), x + f(x) + f(x + f(x)), \ldots$, i.e. $a_0 := x$ and $a_{n+1} = a_n + f(a_n)$. If $a_n < y$ for each n then $(a_n)_{n\in\omega}$ must clearly converge as it is a bounded monotonic sequence. Let a_∞ be the point it converges to. I claim that $f(a_\infty) = 0$, contradicting the assumption that $f(x) > 0$ for any x. By condition (A) on v-frames we know that $|f(a_\infty) - f(a_n)| \leq a_\infty - a_n (= d(a_\infty, a_n))$ for each n. However, since the right hand side converges to 0 as n increases, and $f(a_n)$ converges to 0, it follows that $f(a_\infty) = 0$.

Once one moves away from the simple toy example, v-frames based on $\mathbb{R}^n$ become implausible for other reasons. For example, suppose now we are considering a language in which the only atomic sentences are of the form 'a is red' and 'a is orange' for a a colour in a fixed spectrum. Modelling this language using $\mathbb{R}^2$ would be overly simplistic because the interpretation of 'red' and 'orange' are not independent. Any assignment of cutoff points that allowed 'red' and 'orange' to overlap should be intuitively very far away from the intended interpretation. Thus an interpretation that says that the red colours end at the colour representend by 10 and orange starts at the colour represented by 9 should be very far away from the sensible interpretation that says red ends at 9 and orange starts at 10. However their distance according to the standard metric on $\mathbb{R}^2$ is relatively small: $\sqrt{2} = \sqrt{(10-9)^2 + (9-10)^2}$.

A final worry in this ballpark is that even if $\mathbb{R}^n$ v-frames aren't suitable for modelling vagueness, the correct models might still have enough $\mathbb{R}^n$-like properties to guarantee that the backtrack properties hold. Let me finish by considering two such properties we might quite plausibly expect to hold in any realistic model:

- Density: for any x and y there is a z such that $d(x, z), d(y, z) < d(x, y)$.
- Closeness: Whenever $d(x, y) \leq f(x)$ there is a z such that $d(y, z) \leq f(y)$ and $d(x, z) < d(x, y)$.

However, neither of these principles, even in tandem, ensure that the relevant v-frame has the backtrack property. A simple example would be to let $W := [0, 1) \cup (2, 3]$, $d(x, y) = |x - y|$, $f(x) = 1.5$ for $x \in [0, 1)$ and $f(x) = 1$ for $x \in (2, 3]$. Anything in the range $(\frac{1}{2}, 1)$ can see points in $(2, 3]$, but there is no path from a point in $(2, 3]$ to a point in $[0, 1)$. Furthermore this v-frame is both dense and close. These examples are *gappy*: for a given point x, there may be a range of real numbers $[\alpha, \beta]$, such that the distance between x an another point is never in $[\alpha, \beta]$.

4.2.2 Precise ranges

When considering the simple model presented in this section one might object that it is no better to say that determinately 7 or 8 is the last determinately* small number (although it's indeterminate which) than to say that determinately 7 is the last determinately* small number. Indeed, it seems just as bad to say that the location of this cut-off point is vaguely located over a precise range as it is to say that it is precisely located somewhere.

One might question the intuition that this is just as bad. After all, if we consider the vague predicate 'Small for a number less than 10' it seems fair enough to say that, determinately, either 3 or 4 is the last small number less that 10, although it's indeterminate which. 'small for a number less than 3' seems to be precise.

However there is no need to challenge the intuition. For according to that model there is no precise range in which the last determinately* small number vaguely falls. At the leftmost and middle point the bottom node can see it's vague whether the last determinately* small number is 8 or 9, and at the rightmost point it's vague whether it's 7 or 8. Indeed this is not just a peculiarity of our model. The contrapositive of $(+\Delta)$ says $\neg\Delta\Delta^* p \to \neg\Delta^* p$ so

can show fairly easily that $\nabla\Delta^* p \rightarrow \nabla\nabla\Delta^* p$.[10] In other words, if it's vague whether p is determinate* then it's vaguely vague.

How much of a limitation does this place on our solution? Must we retract all our claims about there being no sharp cut-off points for 'determinately* small' - must we retract them not because they are false, but because they are vague? Evidently we must retract specific claims of the form 'it is vague whether p is determinate*' since they are all at best vague. I claim this is an advantage of the theory, even, since we can avoid the objection that we must have a precise range in which the last determinately* small number falls.

What about the crucial claim that it is vague where the last determinately* small number lies? This claim still stands, and is determinate. In our model not only is it vague, at the bottom node, where the last determinate* number lies, but it's determinately vague where the last determinate* small number lies. This should not be puzzling for the classicist - it is standard to allow it to be determinate that there are F's whilst denying the existence of a determinate F. It is no different for the complex property of being vaguely determinately* small.

4.2.3 The Forced March Sorites and Assertion

One of the more puzzling issues relating to vagueness and higher order vagueness is the so-called 'forced Sorites march.' One is to imagine that you are to be presented with the elements of a Sorites sequence for F in succession, and in each case you are required to say to the best of your ability whether the element is F or not. The puzzle is that there must surely be a first element at which you stop saying 'yes, it's F' and switch to doing something else. Perhaps that is saying 'no, it's not F', or saying 'I don't know' or perhaps it is not saying anything at all. The point is, whatever one does, it seems one is committed to a sharp boundary. (For simplicity I shall confine attention to the responses: 'yes', 'no', 'I don't know', 'it's indeterminate' and saying nothing at all. Other responses are available: 'I'm not sure whether I know or not', 'I don't know whether it's determinate', 'it's indeterminate whether it's indeterminate', and so on, but since these iterations get less and less relevant to the question you're being asked to answer, silence is often a better response even if these answers are known.)

The notion of commitment here is a pragmatic one: which propositions are assertable, i.e. which answers are appropriate in a forced march, depends on the truth of the commitments of an assertion of that proposition. An assertion that p commits you to q iff, had you been as knowledgeable as possible, your assertion would have been appropriate if and only if q.[11] The view I defend states that the strongest proposition an assertion that p commits you to is the proposition that it is determinate that p. Provided you're as knowledgeable as you can be about the relevant facts, p is assertable (ignoring other pragmatic factors) when p is determinate. Thus, for example, saying 'yes, a is F' to a given question commits you to a being determinately F – you should not say this if you think it's vague whether a is F. If you say 'it's vague' then you are committed to its being determinate that it's vague whether a is F, and so on. Not saying anything is importantly different from saying 'I don't know', for the latter commits you to the determinacy of your not knowing, which (I shall argue) entails that you have the ability to know that you don't know. If you don't know whether you know something you are better of staying quiet than saying 'I don't know.'

[10] Proof: let q be $\Delta^* p$, so that we have $\neg\Delta q \rightarrow \neg q$. So $\nabla q \rightarrow (\neg q \wedge \nabla q)$ by definitions and prositional logic. Since we can prove the consequent is not determinate, we can prove in K that the antecedent is not determinate, i.e. we can prove $\neg\Delta\nabla q$, so $\nabla q \rightarrow \neg\Delta\nabla q$. Since $\nabla q \rightarrow \neg\Delta\neg\nabla q$ we have $\nabla q \rightarrow \nabla\nabla q$. Indeed, given the equivalence between q and $\Delta^n q$, it is possible to show that $\nabla q \rightarrow \nabla\Delta^n q$ for any n whatsoever.

[11] If you are not as knowledgeable as you could be then we have only the principle that your assertion is appropriate *only if* its commitments are true. It is only if you have the relevant background knowledge that the truth of the assertions commitments guarantees the appropriateness of the assertion as well.

My case revolves around the principles like the following

> If it's not vague whether Harry is bald, i.e. if there's a fact of the matter whether Harry is bald, then one can in principle know whether Harry is bald. (4.7)

Why might someone object to this principle? Of course, the *general* principle that all precise propositions have knowable truth values might fail for reasons not relating to vagueness at all. But I take it that if the above instance fails at all it fails for reasons relating to the vagueness of 'bald'.

I espouse the principle. According to an opposing view it might be impossible to find out whether Harry is bald, even if it is a determinate matter whether he's bald. According to classical treatments of vagueness it is possible for a proposition to be both determinately true, and for it to be vague whether the proposition is determinately true. In these cases, the opposing view claims, it is not possible to know whether p, *even though* p is determinately true. In these cases the obstacle to our knowing is not vagueness (the proposition in question is determinately true) but second order vagueness. It is very natural, on this view, to also count higher orders of vagueness as sources of ignorance in the same way.[12]

The view that vagueness and only vagueness is responsible for ignorance gives a more intuitive explanation of the phenomenon. The easiest way to explain to a non-philosopher the philosophical notion of vagueness is by saying it's whatever prevents us from knowing whether a given person is bald when we know all the relevant facts about hair number and distribution. Clearly the kinds of reasons we can't know whether a given person is bald or not, even if we know all the facts about hair number and distribution, forms *some* sort of natural kind. The simplest theory simply identifies vagueness with this obstacle to knowledge.[13] As we proceed, we shall see that the simple view does a much better job at explaining the puzzles of higher order vagueness.

The converse of (4.7) seems to be relatively uncontroversial, so given what I have said so far it is reasonable to assume that both (4.7) and its converse are determinate – in other words the following principle is determinate: one can in principle know whether Harry is bald if and only if it's a determinate matter whether he is bald.[14] It is a consequence of this, in the standard logic of determinacy (which includes the K principle), that if it's vague whether Harry is determinately bald, it's vague whether we can know that he's bald. Assuming that the respondent knows as much as she can about Harry's head, it follows that it's vague whether she in fact knows that he's bald, and it thus is also presumably vague whether it's permissible to assert that Harry is bald in such circumstances. None of this should be intrinsically surprising given the inherent vagueness of the relevant notions of possibility, knowledge, permission and assertion.[15] In summary: in a forced march, if you know as much as you can about the situation then 'it's determinate that p', 'you know that p' and 'it's assertable that p' are all coextensive operators.

The companion to this view about knowledge is the view that the strongest proposition an assertion that p commits you to is the proposition that p is determinate. If one is as knowledgeable as possible, then one knows p iff p is determinate. Since, ignoring other pragmatic factors, an assertion is appropriate iff it's knowledgeable, it follows that an assertion that p commits you to no more than the claim that p is determinate in a forced

[12] An explicit example of this view can be found in Fine [44].

[13] One might prefer to simply introduce a separate name for whatever this obstacle to knowledge is: schmagueness. Arguments analogous to those in this chapter would demonstrate that schmagueness iterates non-trivially, and strictly analogous puzzles for higher order schmagueness arise as for vagueness.

[14] I am using 'can' here in a way that guarantees that you cannot know that p if p is false. 'can' cannot simply mean metaphysical possibility here because many false propositions are true, and even known, in some metaphysically possible world.

[15] I take it that the same Sorites sequence that shows that 'x is bald' is vague, shows that 'it's possible to know that x is bald' and 'it's appropriate to assert that x is bald' are vague. I am further claiming that the vague instances of the latter two predicates are precisely the second order vague instances of the former, and the determinate instances of the latter are the determinately determinate instances of the former.

march. In contrast, the companion to the opposing view about knowledge has it that the strongest proposition an assertion that p commits you to is the proposition that p is determinate*.

Let us apply this theory of commitment to the forced march Sorites. We may consider the possible responses in turn. Certainly saying 'yes' up to a certain point, and then saying 'no' commits one to sharp boundaries, for that is to commit one to the elements being determinate cases up to a certain point, and determinate non-cases from there on. This is just what it is to say that F is sharp. To respond by saying 'yes' up to a certain point, a_n, and then continue by saying 'it's indeterminate' or 'I don't know', is slightly more complicated. This commits one to the determinacy of $Fa_1, \ldots Fa_n$. Asserting that it's indeterminate that each of $a_{n+1} \ldots a_m$ is F commits one to the determinacy of $\nabla Fa_{n+1} \ldots \nabla Fa_m$. Thus we are committed to the following: ΔFa_n and $\Delta\neg\Delta Fa_{n+1}$. This is not a commitment to sharp boundaries, but in a standard Sorites (without gaps) one would not expect there to be a determinate F adjacent to a determinately not determinate F. It is thus plausible to assume that this response pattern commits you to some falsehoods; this this pattern of assertions is inappropriate. Finally, if we assume the respondent knows as much as she can about the relevant background precise facts, then her knowledge is coextensive with what's determinate, so if she answers 'I don't know', we may reason as above.

On the other hand, saying 'yes' up to a certain point, and then saying nothing for a stretch is a different matter altogether. This response pattern does not commit one to sharpness of any kind. Not saying anything is not the same as asserting that you don't know Fa_n, since the latter is not appropriate when you do not know that you don't know that Fa_n. Saying nothing does not commit you to any claim about the vagueness or nth order vagueness of Fa_n. It is completely compatible with this response pattern that every predicate of the form $\Delta^n Fx$ has borderline cases. In other words, asserting that the cases are F up to a certain point, remaining silent for a while, asserting the cases are indeterminate, sliding back into silence for a bit, and then asserting the cases are not F from then on, does not commit you to any thing inconsistent with the thesis that $\Delta^n Fx$ is vague for each n.

To demonstrate this, suppose that $Fa_1 \ldots Fa_n$ are determinate, and that $Fa_{n+1} \ldots Fa_m$ aren't. Since there is second order vagueness, it is vague which number n is in our example, but we may be certain there is *some* such n by classical logic. If I happen to say 'yes' from cases a_1 to a_n and remain silent for cases $a_{n+1} \ldots a_m$ I have (a) asserted correctly, in the sense that I have asserted p when p is determinate and (b) have committed my self to nothing incompatible with vagueness at all orders. Furthermore, despite the fact that a_n is the last determinate F, it is presumably vague that a_n is the last determinate F, so my assertions, despite being correct, fail to be determinately correct. If a perfect asserter is someone who asserts p just in case it's determinate that p, there can be perfect asserters but it will always be at best vague whether you're a perfect asserter (provided we assume that it is always a determinate matter whether you have asserted p or not.[16])

It is worth remarking that if you knew you were a perfect asserter, i.e. if you knew that you asserted p just in case it's determinate that p, you would be able to infer from your having not asserted Fa_{n+1} that Fa_{n+1} was not determinate. Since in a typical forced march Sorites, it is at best vague whether you are a perfect asserter, such knowledge would not be available to you. So while there always is a correct response to the forced march Sorites, second order vagueness makes it impossible to know if you've made the correct responses. This should not, in itself be surprising. The maxim 'assert p only if it's determinate that p' is an externalist norm in the sense that one is not always in a position to know whether one is complying with it. Sometimes it is impossible to know whether you are complying because it is impossible to know whether it's determinate that p due to second order indeterminacy.

[16]This assumption may not be unassailable. For example, I might falter or hesitate as I say 'yes' in such a way as to make it vague whether I actually committed myself to the Fness of the case in question. This might be one way to be a determinate perfect asserter.

4.2.4 Other paradoxes

There are a number of other paradoxes of higher order vagueness in the literature which do not rely on the principle B which I shall turn to now. They are both variations on an argument originally due to Wright [146]. We shall see that both these arguments rely on the view about commitment, assertion and knowledge which we have rejected in the previous section. Once the alternative is taken into account it is seen that it is compatible with these results that the thesis that there is vagueness at all orders is both assertable and known.

Let me begin with an argument due to Delia Graff Fara [33]. Fara's argument shows that natural principles concerning higher order vagueness, so called 'gap principles' $\Delta\Delta^n Fa_k \rightarrow \neg\Delta\neg\Delta^n Fa_{k+1}$, with the rule of proof Δ-intro:

If $\Gamma \vdash \phi$ then $\Gamma \vdash \Delta\phi$

lead to a contradiction. My focus here will be the rule of proof of Δ-introduction.

There has been some debate concerning whether a supervaluationist should accept classical logic, where 'classical logic' is construed broadly to include classical rules of inference and rules of proof. The rule Fara appeals to is incompatible with certain classical rules of proof. For example one could not apply conditional proof to $p \vdash \Delta p$ (which can be obtained by Δ intro on $p \vdash p$) to obtain $\vdash p \rightarrow \Delta p$, since this would imply, as a matter of logic, that everything is precise.

There are some tricky questions in this area concerning the nature of logic, whether Δ is a logical constant, and on the normative impact both notions of consequence have. Since Fara's conclusion is essentially normative - that it is in some sense incoherent to accept vagueness at all orders - it would be nice to talk directly about the normative conclusions without the detour through 'consequence' talk which can be quite obscure in these contexts anyway. Let me introduce two notions of a 'good inference' from $p_1 \ldots p_n$ to q.

1. $Cr(q) = 1$ if $Cr(p_i) = 1$ for each $i \leq n$ and $Cr \in \mathsf{E}$.
2. $\Sigma_{i \leq n}(1 - Cr(p_i)) \leq 1 - Cr(q)$ for every $Cr \in \mathsf{E}$.

Here E is a set of credence functions that you would be justified in having in some possible epistemic situation. One notion governs what can be inferred given what you are already fully justified in believing, whereas the other constrains your beliefs when you are less than certain in the premisses. Something like these two notions are sometimes characterised in terms of global and local consequence, corresponding to the definiteness of the premisses strictly implying the definiteness of the conclusion (i.e. 'preservation of supertruth') and the premisses simply strictly implying their conclusion (i.e. 'preservation of disquotional truth'.) In formal terms that is $\Box(\Delta p_1 \wedge \ldots \wedge \Delta p_n \rightarrow \Delta q)$ and $\Box(p_1 \wedge \ldots \wedge p_n \rightarrow q)$ where $\Box$ represents some suitable notion of logical necessity.

Since one could never find oneself in an epistemic situation which *fully* supported $p \wedge \neg\Delta p$, but one could quite easily have evidence for $\neg p \vee \Delta p$ the former notion of good inference invalidates reductio: we have $p \wedge \neg\Delta p \vdash$ but not $\vdash \neg(p \wedge \neg\Delta p)$.

I am happy to engage in either talk provided it is clear what one means and one is careful which normative conclusions one draws. However, neither notion permits the inference from p to Δp. Crucially the first notion, 1., does not permit this inference. Observe first that this rule does not preserve determinate truth, for example if p is determinate but not determinately determinate, the premise of $p \vdash \Delta p$ is determinate and its conclusion isn't. Similarly, since p is precise (although not determinately so), one could in principle have evidence which justifies certainty in p, yet be uncertain in Δp due to the vagueness in Δp. To be sure this counterexample relied on higher order vagueness, but this is clearly not the place to beg *that* question.

Note that the consequence relation that instead preserves determinacy* does appear to validate Δ-intro. For example, according to this relation $(+\Delta)$ guarantees that $p \vdash \Delta p$. As I have argued in the previous section, if your beliefs are 'inconsistent' according to this

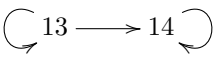
consequence relation you are not necessarily being incoherent by committing yourself to a contradiction. The most any belief or assertion commits you to is the determinacy, not the determinacy*, of what is believed or asserted.

In this vein Zardini presents an argument that does not rely on Δ-intro [152]. His argument operates directly with the notion of determinacy*. His argument shows that if we assume the determinacy* of (a) the vagueness of $\Delta^n Fx$ for each n, (b) the Fness of a_0 and (c) the non-Fness of $a_{1,000,000}$ we can derive a contradiction. More formally he assumes the following: $\Delta^*\exists x\nabla\Delta^n Fx$, $\Delta^* Fa_0$, $\Delta^*\neg Fa_{1,000,000}$.

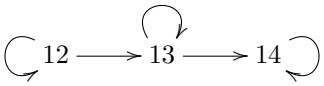
What is surprising about Zardini's argument is that although I have throughout been arguing for the vagueness of predicates of the form $\Delta^n Fx$, i.e. for the claim $\exists x\nabla\Delta^n Fx$, I cannot maintain that this claim is determinate*. This is interesting as it is a concrete example of something I would count as a permissible assertion which is not determinate*.

To check that these assertions, including $\exists x\nabla\Delta^n Fx$ and $\exists x\nabla\Delta^* Fx$, do not commit us to any contradictions we need to show that not only are there models in which they are all true, but that there are models in which they are all determinately true. In fact, one can show for any $n \in \omega$ that there is a model in which these claims are determinate^n true.

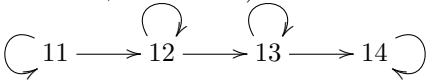
Let's start with an example in which $\exists x\nabla\Delta^* Fx$ is simply true. Note also that all these examples apply also to $\exists x\nabla\Delta^n Fx$.



Remember that a number is determinately* small at a point iff you can't get to a smaller number by following the arrows. So, for example, the left node sees a node (itself) in which 14 is not determinately* small, and can see a node in which it is (the right node.) Thus, at the left node it is vague whether 14 is determinately* small, so 'determinately* small' has a borderline case. However the vagueness of 'determinately* small' is not determinate because the left node sees a node in which 'determinately* small' is completely precise: the right node.



Here at the leftmost node it is vague whether 13 is determinately* small: it sees a world where it isn't (itself) and a world where it is (the middle node.) At the middle node it's vague whether 14 is determinately* small (see above). So at the leftmost node we have $\Delta(\nabla\Delta^* S(13) \vee \nabla\Delta^* S(14))$. Thus at every node the leftmost node sees 'determinately* small' has a borderline case, witnessed by 13 and 14 respectively. This gives a model for $\Delta\exists x\nabla\Delta^* Fx$ (and also $\Delta\exists x\nabla\Delta^n Fx$, for each n.)



Just as before, it is vague at the leftmost node whether 12 is determinately* small. At every world it sees either 12 or 13 is a borderline case of determinate* smallness (see above), and at every node seen by a node seen by the leftmost either 12, 13 or 14 is a borderline case of determinate* smallness. So we have $\Delta\Delta(\nabla\Delta^* S(12) \vee \nabla\Delta^* S(13) \vee \nabla\Delta^* S(14))$. Thus this is a model for $\Delta\Delta\exists x\nabla\Delta^* Fx$ (and also $\Delta\Delta\exists x\nabla\Delta^n Fx$, for each n.) It should be clear how to carry on this series.

4.2.5 Nihilism*

One response to a Sorites paradox for a predicate of the form $\Delta^* Fx$ is simply to deny than *anything* satisfies $\Delta^* Fx$. I shall call this view 'nihilism*' since it mimics the nihilist response to the ordinary Sorites paradox.

It should be noted that the necessitation principle for Δ, a principle we have appealed to throughout this chapter, already supplies counterexamples to nihilism* since, with a

principle of infinitary conjunction introduction, we can easily show $\Delta^*(Fx \vee \neg Fx)$. Not only can we show that logical truths are determinate*, but we can also show things like $\Delta^*(\Delta Fx \to Fx)$, and indeed $\Delta^*(\Box Fx \to Fx)$ and $\Delta^*(KFx \to Fx)$ if we have necessity and knowledge operators in the language. Although certain logical and conceptual truths are among the determinate* propositions, some conceptual truths aren't. For example, if it is vague whether any foetus of a certain age is a person, then presumably it is is a conceptual truth one way or the other without being a determinate, and hence a determinate*, truth one way or the other.

Anyone who accepts necessitation for Δ has to admit a distinction between certain determinate* truths and others. I would be very skeptical that such a distinction would be a precise distinction.[17] And without a motivated precise distinction between the Δ^* truths and the rest, a response to the paradoxes of higher order vagueness is needed.

A more radical approach is to deny necessitation altogether. This is the strategy that Dorr seems to endorse in [26]. A standard way to model failures of necessitation is to introduce a distinction between normal and non-normal nodes in the kind of Kripke frames we have been considering. Such a move invites variants of the kinds of paradoxes we have been considering. For example, we could define an operator $\Delta_N p$ saying that p is true at all accessible *normal* worlds, and run the paradox for the operator $\Delta^*_N p$. Without appealing to this particular semantics for non-necessitatable operators, we must have some notion of normalcy which allows us to say things like '$\Delta p \to p$ is part of our logic and $\Delta p \to \Delta\Delta p$ isn't', i.e. something stronger than the mere truth or assertability of p, but weaker than the empty notion of determinate* truth.

4.3 Appendix

Here we demonstrate some relevant facts about the logic of higher order vagueness. Recall that:

Definition 4.3.0.2. *A v-frame is a triple $\langle W, d(\cdot,\cdot), f(\cdot)\rangle$ where $\langle W, d\rangle$ is a metric space, and $f : W \to \mathbb{R}^+$ obeys the following:*

(A) $\forall w, v \in W, |f(w) - f(v)| \leq d(w, v)$

A formula of propositional modal logic is valid on a v-frame $\langle W, d, f\rangle$ iff it is valid on the Kripke frame $\langle W, R\rangle$ where Rxy iff $d(x, y) \leq f(x)$.

Dorr [25] shows, translating into the terminology of v-frames, that B is not valid over the v-frame $\langle (1, 2), |x - y|, \frac{x}{3}\rangle$ although the weaker principles $p \to \Delta\neg\Delta\Delta\neg p$ and B^2 are valid in this frame. It is possible, however, to construct v-frames in which $p \to \Delta\neg\Delta^n\neg p$ is valid for no $n \in \mathbb{N}$. For example, let $W := \{0, 1\}$, $d(x, y) = |x - y|$, $f(0) = 1$ and $f(1) = \frac{1}{2}$.

What is the logic of v-frames? Clearly every v-frame generates a corresponding reflexive Kripke frame, so the logic of v-frames contains KT. One might have hoped that every reflexive Kripke frame could be generated from a v-frame this way ensuring a logic of exactly KT. This reduces to the question of whether every reflexive digraph can be embedded into a metric space in such a way that there is a a closed ball around each node that contains all and only those nodes it can see. Unfortunately this does not hold:

Fact: Suppose $\mathcal{F}$ is a Kripke frame based on a v-frame. If $\mathcal{F}$ contains a cycle, it contains a 2-cycle.

To see this suppose that $\langle a_0, \ldots, a_n\rangle$ is a cycle in $\mathcal{F} = \langle W, R\rangle$ where $n > 2$. For convenience let $a_i = a_j$ where $j = i \bmod (n + 1)$ for $i > n$. Now suppose that that $\neg Ra_{i+1}a_i$ for every i. Since for each i Ra_ia_{i+1} we know that $d(a_i, a_{i+1}) \leq f(a_i)$ in the corresponding v-frame. We also know that $f(a_i) < d(a_{i-1}, a_i)$ since $\neg Ra_ia_{i-1}$. Thus for each i, $d(a_i, a_{i+1}) \leq f(a_i) < d(a_{i-1}, a_i)$, so $f(a_n) < d(a_{n-1}, a_n) \leq f(a_{n-1}) < \ldots \leq$

[17] At least, there isn't any obvious precise criteria for distinguishing the two such as 'is a tautology' and so on.

$f(a_{-1}) = f(a_n)$, i.e. $f(a_n) < f(a_n)$ which is a contradiction. So for some i, Ra_ia_{i+1} and $Ra_{i+1}Ra_i$.

v-frames thus have more structure than reflexive frames. However, it turns out this does not make a difference to the logic:

Theorem 4.3.1. Completeness. *A set Σ is valid on every v-frame iff its members are theorem's of* KT.

Proof. Suppose that Σ is a KT-consistent set of formulae. Then Σ is satisfiable on the canonical frame $\mathcal{F}$. $\mathcal{F}$ may contain cycles without 2-cycles, so we cannot yet infer that Σ is satisfiable on some v-frame. However we may construct a frame from $\mathcal{F}$, with all the cycles ironed out, that is equivalent to a v-frame.

Let a_0 be a maximal KT-consistent set containing Σ. We may assume that a_0 is a root of $\mathcal{F}$ (if it isn't take the generated subframe around a_0 and work with that instead.) Define $\mathcal{F}^+ := \langle W^+, R^+\rangle$ as follows

- $W^+ := \{s \mid s \text{ a path in } \mathcal{F} \text{ such that } s_0 = a_0\}$
- $R^+ := \{\langle s, t\rangle \mid |t| = |s| + 1 \text{ and } s_i = t_i \text{ for } i \leq |s| \text{ or } s = t\}$

Claim: $f(a_0, \ldots, a_n) = a_n$ is a bounded morphism from $\mathcal{F}^+$ to $\mathcal{F}$.

(1) Suppose R^+st. If $s = t$ then $f(s) = f(t)$ so $Rf(s)f(t)$ since R is reflexive. If $|t| = |s| + 1$ then $Rf(s)f(t)$ since t and s are paths.

(2) Suppose $Rf(s)f(t)$. We want to find a u such that R^+su and $f(u) = f(t)$. If $f(t) = f(s)$ let $u = s$. Otherwise, let $u = \langle s, f(t)\rangle$.

Since anything valid on $\mathcal{F}^+$ is valid on every bounded morphic image of $\mathcal{F}^+$ (see for example [10]) it follows that Σ is satisfiable on $\mathcal{F}^+$. Now we construct our v-frame as follows:

- We begin by defining distance between adjacent points. If R^+st then $e(s,t) = e(t,s) = \frac{1}{2^{|s|}}$. Always fix $e(s,s) = 0$
- $d(s,t) := inf\{\sum_{i=0}^{n} e(p_i, p_{i+1}) \mid p_0 = s, p_n = t, p \text{ a path in the symmetric closure of } \mathcal{F}^+\}$
- $f(s) := \frac{1}{2^{|s|}}$

It is now easy to check that $\langle W^+, d, f\rangle$ is a v-frame and that R^+st iff $d(s,t) \leq f(s)$. □

Cian Dorr has pointed out to me that the constraint (A) on v-frames does not play much of a role in the proof of Theorem 4.1. This allows us to prove a slightly more general result:

Definition 4.3.1.1. *a* **difference measure** *is a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that:*

- *g is continuous in both arguments.*
- $g(x,x) = 0$
- $g(x,y) = g(y,x)$ *(this constraint is optional in what follows.)*

For a given difference measure, g, a g-frame is a triple $\langle W, d(\cdot,\cdot), f(\cdot)\rangle$ where $\langle W, d\rangle$ is a metric space, and $f : W \to \mathbb{R}$ such that:

(A′) $\forall w, v \in W, g(f(w), f(v)) \leq d(w,v)$

Corollary 4.3.2. *For any difference measure g, the logic of g-frames is* KT.

Proof. Note that for any positive a there is a $b < a$ such that $g(a,b) \leq a$ since $g(a,a) = 0$ and g is continuous in both arguments. For any a pick a unique such b, a_g (choice.)

Now modify the construction in Theorem 4.1 as follows.

- Fix $e(s,s) = 0$ for every s.
- Let $e(\langle a_0\rangle), t) = e(t, \langle a_0\rangle) := 1$ for $t \neq \langle a_0\rangle$ such that $R^+\langle a_0\rangle, t$.
- Suppose that $e(s,t) = e(t,s) = a$ has already been defined for R^+st, and suppose that R^+tu $t \neq u$. Define $e(t,u) = e(u,t) = a_g$.
- $d(s,t) := inf\{\sum_{i=0}^{n} e(p_i, p_{i+1}) \mid p_0 = s, p_n = t, p \text{ a path in the symmetric closure of } \mathcal{F}^+\}$
- $f(s) := sup\{e(s,t) \mid R^+st\}$.

□

The interest in this generalization is that one might think that it becomes much harder for two points to differ on the interpretation of 'determinately' the closer together they are. Perhaps it is not the difference between $f(w)$ and $f(v)$ that must be less than $d(w,v)$ but the difference between their ratios, or some other such g.

4.3.1 Further restrictions

The proof of completeness and the counterexamples to the B^n principles relied heavily on our considering slightly artificial frames that were based on metric spaces that either aren't dense, or have points with zero accessibility range. A natural class of v-frames to consider are those based on metric spaces of the form $\mathbb{R}^n$ where $f(a) > 0$ for all $a \in \mathbb{R}^n$. In these frames whenever x can see y, there is a path back from y to x, even though there are frames invalidating B^n for each $n \in \omega$ (i.e. there is no upperbound on how long these paths might be.) This is worrying since this means that $\Delta\Delta^*p \vee \Delta\neg\Delta^*p$ is valid over these frames.

We can express something like this principle in modal logic. I'll call it B^*.

$$\mathsf{B}^*\text{: } \Delta(p \to \Delta p) \to (\neg p \to \Delta\neg p) \tag{4.8}$$

B^* is valid in the class of v-frames just described. In the presence of KT, B^* defines what I shall call 'the backtrack principle'.

> Whenever Rxy there exists $z_1, \ldots, z_n$ such that (a) $z_1 = y$, $z_n = x$ and Rz_iz_{i+1} for $1 \leq i < n$ and (b) Rxz_i for $1 \leq i \leq n$. (4.9)

Proof. We shall show that $(\Delta(p \to \Delta p) \wedge \neg\Delta\neg p) \to p$ defines the requisite property. Suppose the reflexive frame $\mathcal{F} = \langle W, R\rangle$ has the backtrack property. Now suppose $x \Vdash (\Delta(p \to \Delta p) \wedge \neg\Delta\neg p)$. The second conjunct ensures that there is a y such that Rxy and $y \Vdash p$. Since $\mathcal{F}$ has the backtrack property there is a finite path back from y to x, $z_1, \ldots, z_n$, which x can see. Since $x \Vdash \Delta(p \to \Delta p)$ each $z_i \Vdash p \to \Delta p$. Since $z_1 = y$ and $y \Vdash p$, $y \Vdash \Delta p$ - by induction we can see that $z_i \Vdash p$ for each i which means $z_n = x \Vdash p$ as required.

For the other direction suppose, for contradiction, that $\mathcal{F} \models (\Delta(p \to \Delta p) \wedge \neg\Delta\neg p) \to p$ but $\mathcal{F}$ lacks the backtrack property. This means that for some x and y, Rxy but there is no path back from y to x which x can see. Define the following valuation on $\mathcal{F}$: $w \Vdash p$ iff there are $z_1, \ldots z_n$ such that (1) $z_1 = y$, Rz_nw and Rz_iz_{i+1} for $1 \leq i < n$ and (2) Rxz_i for $1 \leq i \leq n$. Certainly if x had this property then $z_1, \ldots, z_n, x$ would be a path back to x which x can see, so $x \nVdash p$. However $x \Vdash \Delta(p \to \Delta p)$ since if Rxw and $w \Vdash p$ then there is a path from y to w satisfying (1) and (2): $z_1, \ldots z_n$. Furthermore, for any world that w sees, w', $z_1, \ldots, z_n, w$ will be a path from y to w' satisfying (1) and (2), since Rxw.

□

Is KTB^* the modal logic of these v-frames? We start with a negative result: KTB^* is not sound and *strongly* complete with respect to *any* class of frames. I.e. there is no class of frames, $\mathcal{C}$, such that a set is KTB^* consistent iff it's satisfiable on a frame in $\mathcal{C}$. The following also shows it is neither canonical nor compact.

Proof. To show this we shall show there is a KTB^*-consistent set of sentences which is unsatisfiable on every frame validating KTB^*.

Let $\Sigma := \{p, \neg\Delta\neg q\} \cup \{\Delta(q \to \Delta^n \neg p) \mid n \in \omega\}$. If Σ were KTB^*-inconsistent some finite subset would be KTB^*-inconsistent (since proofs are finite.) We shall show that for every $m \in \omega$, $\Sigma_m := \{p, \neg\Delta\neg q\} \cup \{\Delta(q \to \Delta^n \neg p) \mid n \in m\}$ is KTB^*-consistent. Σ_m has a KTB^*-model: $\langle m+1, R\rangle$ where Rxy iff $x = 0$ or $x > 0$ and $|x - y| \leq 1$. 0 can see m and there is a finite m length path back from m to 0 that 0 can see but no shorter path. Let q be true only at m and p only at 0.

However, if $\mathcal{F}$ validates KTB^* then $\mathcal{F}$ has the backtrack property so at no point of $\mathcal{F}$ is every member of Σ true: if $x \Vdash \neg\Delta\neg q$ then x sees some $y \Vdash q$. By the backtrack property there is a path $z_1, \ldots, z_n$ back to x which x can see, so $\Delta(q \to \Delta^{n+1}\neg p)$ cannot be true at x if $x \Vdash p$.

□

However there is a positive result, namely that KTB^* is sound and complete over the class of reflexive frames with the backtrack property. For this result I refer the reader to Benton [9], who shows that KTB^* has the finite model property.

Theorem 4.3.3. *If ϕ is KTB^*-consistent then it is satisfiable on a finite reflexive frame with the backtrack property.*

The question whether KTB^* the logic of v-frames over $\mathbb{R}^n$ in which $f(x) > 0$ remains open.

4.3.2 B entails that a conjunction of determinate truths is determinate

The aim is to show distributivity within the modal logic KB with infinitary conjunction (C1-C3 below.) For convenience I shall introduce an operator $\Diamond p := \neg\Delta\neg p$

D. $\bigwedge_{i<\omega} \Delta\phi_i \to \Delta \bigwedge_{i<\omega} \phi_i$.

KB

K $\Delta(\phi \to \psi) \to (\Delta\phi \to \Delta\psi)$

B $\phi \to \Delta\Diamond\phi$

Nec. if $\vdash \phi$ then $\vdash \Delta\phi$

C1. $\bigwedge_{i<\omega} \phi_i \to \phi_n$ for each $n < \omega$.

C2. $\bigwedge_{i<\omega}(\phi_i \to \psi_i) \to (\bigwedge_{i<\omega} \phi_i \to \bigwedge_{i<\omega} \psi_i)$.

C3. If $\vdash \phi_i$ for each $i < \omega$, $\vdash \bigwedge_{i<\omega} \phi$.

Claim: D is independent of K (and KT) + C1-C3.

Construct a Montague-Scott frame as follows (see [19]): Let $\mathcal{W} := \mathbb{N}$ and for each world $w \in \mathcal{W}$ let the necessary propositions at w, $N(w)$, be the cofinite subsets of $\mathcal{W}$ (if we are trying to model T (i.e. $\Delta\phi \to \phi$) as well we let $N(w) := \{X \cup \{w\} \mid X \text{ is cofinite}\}$.) Then $\langle \mathcal{W}, N\rangle$ satisfies:

1. $\mathcal{W} \in N(w)$ for all $w \in \mathcal{W}$
2. If $X, Y \in N(w)$ then $X \cap Y \in N(w)$
3. (For T) If $X \in N(w)$ then $w \in X$.

Thus our frame models K (/KT) including C1-C3. However it does not model D, as can be seen by letting $[\![p_i]\!] := \{n \in \mathbb{N} \mid n > i\}$ (for KT: $\{n \in \mathbb{N} \mid n > i\} \cup \{0\}$, allowing D to fail at 0.) On the other hand any Kripke frame (reflexive Kripke frame) will validate K (KT) along with D.

Although the distributivity of infinite conjunctions over Δ is independent of K, the distributivity of $\Diamond$ over infinite conjunctions, perhaps surprisingly, is not independent in this way and can be show given just some relatively uncontroversial principles governing infinite conjunction.

Lemma 4.3.4. $\Diamond \bigwedge_{i<\omega} p_i \to \bigwedge_{i<\omega} \Diamond p_i$

Proof. First note that $\Diamond \bigwedge_{i<\omega} p_i \to \Diamond p_j$ for each j, by C1 and the background modal logic of K. Then by C3 and then C2 we can infer $\Diamond \bigwedge_{i<\omega} p_i \to \bigwedge_{i<\omega} \Diamond p_i$. □

Theorem 4.3.5. *Although D is independent of K (and KT) it is not independent of, and is in fact entailed by, KB (and thus KTB.)*

Proof. B directly gives us:

$$\bigwedge_{i<\omega} \Delta p_i \to \Delta\Diamond \bigwedge_{i<\omega} \Delta p_i \tag{4.10}$$

We may also infer from our lemma that

$$\Delta\Diamond \bigwedge_{i<\omega} \Delta p_i \to \Delta \bigwedge_{i<\omega} \Diamond\Delta p_i \tag{4.11}$$

by applying necessitation and the K principle. Finally we have

$$\Delta \bigwedge_{i<\omega} \Diamond\Delta p_i \to \Delta \bigwedge_{i<\omega} p_i \tag{4.12}$$

because we have $\Diamond\Delta p_i \to p_i$ for each i by B. So by C3 and then C2 we get $\bigwedge_{i<\omega} \Diamond\Delta p_i \to \bigwedge_{i<\omega} p_i$, and by necessitation and K that gives (12).

But (10), (11) and (12) give distributivity. □

It should be noted that this argument does not appeal to any characteristically classical principles. Indeed this argument can be carried out provided one has the following rules of inference as primitive or derived:

$\to$1. $\phi, \phi \to \psi \vdash \psi$

$\to$2. $\phi \to \psi, \psi \to \chi \vdash \phi \to \chi$

$\to$3. $\phi \to \psi \vdash \neg\psi \to \neg\phi$

and where KB is understood to contain K and the axioms $\phi \to \Delta\Diamond\phi$ and $\Diamond\Delta\phi \to \phi$.[18]

[18]Without the second version of B the most we could prove without double negation elimination would be things of the form $\Diamond\Delta\neg\phi \to \neg\phi$.

Chapter 5

A Theory Of Indeterminacy

In this thesis we have argued against the view that indeterminate and factually defective discourse arises due to a certain kind of laxness in the way we use language to talk about the world. In this chapter we outline an alternative theory of indeterminacy which does not depend on the way we use language.

In doing so we also answer a potential challenge: one might concede that vagueness is not linguistic but still think that *some* indeterminacy is linguistic. Two particular examples come to mind: indeterminacy arising from concepts introduced via incomplete linguistic stipulations and indeterminacy arising when names for the roots of -1 are introduced independently [14]. In §5.1 I outline these two examples and in §5.2 I develop a general theory of indeterminacy which accounts for these examples. Finally, in §5.3 I present some formal machinery which I think is illuminating for the account presented here.

5.1 Two examples of indeterminacy

Many topics of interest to philosophers are thought to involve indeterminacy. A short list of such examples includes:

This sentence is true	(5.1)
If Bizet and Verdi had been compatriots, they would have been Italian.	(5.2)
'This ball on earth is heavier than that ball on the moon' said in pre-Newtonian times.	(5.3)
Harry is bald.	(5.4)
It will rain tomorrow.	(5.5)
There's an uncountable subset of the reals that's not in one-one correspondence with the reals.	(5.6)

While none of these examples is uncontroversial – indeed, for each example some philosopher has disputed its claim to being indeterminate – it seems clear that most philosophers understand what it means to call a claim indeterminate, even if only to dispute it. However, while I think it is relatively clear that some of these examples are somehow factually defective, I don't believe that vocabulary like 'factual', 'indeterminate' or 'no fact of the matter' have clear and pretheoretically available meanings. For example, as many philosophers use it, the claim that p and it is not a fact that p is consistent; yet the pretheoretical use of these terms seems to suggest otherwise. The project, as I understand it, is to elucidate the notion to someone who already has it, or at least to isolate some notion which has some intuitive claim to expressing non-factuality which plays the right theoretical role.

My focus will be two examples that appear to be factually defective, which seem particularly well suited to a linguistic explanation of this defectiveness.

The first example is adapted from Brandom [14] (although I shall follow Field's presentation in [36].) Imagine that at some point before the 16th century a community of English speakers had been separated from the rest of us and that they had then been reintegrated with us at the present day. We find out that their mathematicians had also discovered the complex numbers. Now their dialect of English is much like ours, except that in introducing names for the two square roots of -1 they used the symbols '/' and '\', in stead of 'i' and '$-i$'. We may suppose that the two dialects of English become integrated, but the four names for the two square roots of -1 remain. Now it is easy to show that i and $-i$ are the only two square roots of -1, and since / is a square root of -1, so we can prove that $/ = i$ or $/ = -i$. But which of these two possibilities obtains? The dominant view seems to be that there is no fact of the matter; it is indeterminate whether $/ = i$ or $/ = -i$. This is because there is an automorphism of the complex numbers which maps i to $-i$, so there is no '/' free sentence the isolated community could utter that would distinguish / from i without it also distinguishing it from $-i$.

At any rate, whatever this response amounts to, it presupposes the intelligibility of some notion of indeterminacy or non-factuality. The goal of this chapter is to get a better understanding of this notion. Since the use of an indeterminate identity statement adds a wrinkle which I don't think is important, I shall take my first example of an indeterminate proposition to be:

$$/ \text{ is positive} \tag{5.7}$$

where an imaginary number is taken to be positive if it is a positive real multiple of i, and negative if it is a negative real multiple of i. We assume that 'positive' will not be a notion the separated community will have until they are reintegrated.

My second example is of an incomplete definition, taken from Fine [43]. Suppose that I stipulate that a natural number is nice* if it is less than 15 and that it's not nice* if it is larger than 15. Assuming that this stipulation is good we can go on to say many contentful things about being nice*. For example, we may say that there are nice* primes, or that there are at least 10 nice* numbers. It seems that being nice* is a perfectly legitimate property, which some numbers have and others don't. The difficulty concerns cases like the following:

$$15 \text{ is nice}^* \tag{5.8}$$

What is the status of (5.8)? Assuming our stipulation is good, it follows from the law of excluded middle that either 15 is nice* or it isn't, but there is no way of determining which from our stipulation. Of course, one could question whether the stipulation succeeded, and that our uses of 'nice*' ever express a property. However it is common practice to make use of a word without there being explicit necessary and sufficient conditions for its application in all cases, and the stipulation we made seems similar enough to this phenomenon to make it hard to see how it might differ. For example, even in mathematics, we freely make use of exponentiation even though there are many functions, f, which agree with our use of exponentiation but differ over the value of $f(0,0)$. Although there is slightly less consensus in this case, one of the most prominent accounts of (5.8) is again to say that there no fact of the matter.[1] It is indeterminate whether 15 is nice* or not.

5.2 A non-linguistic account

In what follows I shall outline a non-linguistic account of indeterminacy. Here we bring together the ideas developed in other chapters by defining the notion of determinacy explicitly

[1] Other views include the view that 'nice*' does not express a proposition, ([5]) and the view that 15 is (determinately!) not nice* [140].

within our theory of propositional attitudes; i.e. we define the property of a proposition being determinate in terms of the notion of a coherent prior and preference ranking.

However, an intermediate notion I shall appeal to in my analysis is that of a proposition being *indiscernible* from another, which I think has intuitive appeal of its own. I shall begin by introducing the notion of indiscernibility informally, and I shall use it to give a definition of what it means for a proposition to be determinate or indeterminate. Then in section 5.2.3 I shall show how to reduce the relation of indiscernibility to the notion of a coherent preference ranking and a coherent prior which we introduced in chapter 1.

One might object that we cannot have the concept of conceptual coherence without already understanding the notion of indeterminacy I am trying to introduce. Whether or not my explication constitutes a *reduction* of the concept of indeterminacy to concepts not involving indeterminacy is not a question I shall concern myself with here. Throughout this thesis I have argued that the notion of indeterminacy plays an important role within a theory of rational propositional attitudes. Through this role we have introduced the notion of indeterminacy implicitly; here we show how to introduce it explicitly in terms of the notions of coherent prior and preference. No claim of reducibility should be inferred from this.

5.2.1 Indiscernibility

Recall that / and \ are the two square roots of -1, but it is undetermined whether / is i or $-i$. The fact that there is an automorphism of the complex numbers that maps / to \ ensures that / cannot be structurally distinguished from \. But they are not just indiscernible from the point of view of mathematical relations: every belief or desire that could be rationally held about /, would be equally rational if / were everywhere substituted for \ and vice versa (a similar thing can be said about i and $-i$.) Furthermore, it seems that if you're rational then your beliefs and preferences ought to be invariant under a uniform substitution of / for \ and vice versa.[2] Thus anything of importance to communication, decision making and so on is preserved if we uniformly replace the concept of / with \ in everything we think and say. Indeed, there is nothing we can think or say about / which can't be said also about \ that doesn't essentially involve / or \ in some way. This fact is reflected at the level of propositions too: there an automorphism, this time on the algebra of propositions, that preserves their role in our thought and interaction with the world (I will make this precise later) which switches the proposition that $/ = i$ with the proposition that $\backslash = i$. For all intents and purposes the proposition that $i = /$ with $i = \backslash$ are interchangeable - for any calculation I might want to do, or belief or desire I might form, it wouldn't matter if I had it about the proposition that / is F or the proposition that \ is F so long as I am rational and am consistently interchanging every thought and belief I form in these propositions.

In our second example we stipulated that a number is nice* if it is less than 15 and not nice* if it is greater. Suppose I now introduce a new property, being nice$_*$, as follows:

A number is nice$_*$ iff it is either less than 15, or equal to 15 and not nice*. (5.9)

Notice that a number is nice* if it is less than 15, it is not nice$_*$ if it is greater than 15 and 15 is nice$_*$ if and only if 15 is not nice*. There is a symmetry in these definitions which isn't immediately apparent. For if we already had the notion of being nice$_*$ we could define nice* as follows:

A number is nice* iff it is either less than 15, or equal to 15 and not nice$_*$. (5.10)

[2]This latter claim seems to be stronger, since the former claim leaves it open that there are some rational people who would have a preference involving / but not \, provided there are other rational people who have the reverse preference.

Now we can easily see that the proposition that 15 is nice* is distinct from the proposition that 15 is nice$_*$: the truth of one proposition implies the falsity of the other, so they cannot be identical. Similarly neither proposition is identical to the proposition that 15 is less than or equal 15, or the proposition that 15 is less than or equal to 14, since we presumably don't know whether 15 is nice* or nice$_*$, but we surely know that 15 is less than or equal to 15 but not less than or equal to 14. Nevertheless, that each proposition implies the negation of the other seems to be the only important difference between them. Anything else that we might think or say to distinguish them would involve the new propositions in some way or another (again, I shall defer making this precise until later.) The proposition that 15 is nice* and the proposition that 15 is nice$_*$ can't be distinguished in terms of the role they play in our mental lives. They bear exactly the same evidential relations to other propositions, and cannot be treated asymmetrically by a rational agent within her system of beliefs and preferences.

With the notion of indiscernibility at hand we are already in a position to state an account of indeterminacy. A feature both examples have in common is the existence of propositions indiscernible from them. We can use this to explain what it means for a proposition to be potentially indeterminate or 'totally determinate' (i.e. not potentially indeterminate.)

> It is totally determinate whether p iff there are no other propositions indiscernible from p. (5.11)

We might compare this to the notion of a proposition being potentially borderline: according to this conception a proposition is totally determinate if it is not even 'potentially indeterminate'. The familiar determinacy operator can be obtained in the following way:

> It's determinate that p iff every proposition indiscernible from the proposition that p is true. (5.12)

Thus, for example, the proposition that 15 is nice* is not determinate: it is indiscernible from itself and the proposition that 15 is nice$_*$. And since these two propositions are inconsistent at least one of them must be false. Therefore not every proposition indiscernible from the proposition that 15 is nice* is true.

For an example of a proposition that is determinate but not totally determinate consider the proposition, p, that either there are no flying pigs or 15 is nice* and the proposition, q, that either there are no flying pigs or 15 is nice$_*$. p and q are indiscernible from each other and from themselves, so neither are totally determinate. However they are both true, since there are no flying pigs, so both are determinate.[3]

This also raises a distinction between 'pure' indeterminate propositions, such as the proposition that 15 is nice* and impure, such as the proposition that either there are no flying pigs or 15 is nice*.

> p is a pure indeterminate proposition iff the conjunction of every proposition indiscernible from it is the contradictory proposition, and the disjunction of every proposition indiscernible from it is the tautologous proposition. (5.13)

This brings us to another natural question. It is not clear yet whether (5.12) is adequate. Does (5.12) guarantee that there are people who are indeterminately bald? It is not obvious, yet, that it does. This is partly due to our focus on pure indeterminate propositions, such as the proposition that 15 is nice*, instead of contingently indeterminate propositions (yet even the disjunctive example above does little to make claim about baldness plausible.) It should become clear as we progress, however, that this approach applies to vagueness:

[3]Note also that this is an example of two indiscernible propositions which are not negations of each other.

a proposition, such as the proposition that Harry is bald, can usually be split up into a disjunction of conjunctions of pure indeterminate propositions and totally determinate propositions; at a first approximation, with idealising assumptions, the proposition that Harry is bald is the disjunction of the set {Harry has no more than n hairs and having m hairs is the cut off for baldness | $n \leq m$}.

5.2.2 Logical properties of indiscernibility

The preceding remarks make a reasonable case that there is some intuitive notion of indiscernibility. We now attempt to give a more precise account of the logical structure of the indiscernibility relation.

Let us symbolise the relation of indiscernibility as $\sim$. The claim that p is totally determinate can be thus formalised as $\forall q(p \sim q \rightarrow p = q)$. The following (very incomplete) list of principles look as though they govern the indiscernibility relation:

1. $p \sim p$
2. If $p \sim q$ and $q \sim r$ then $p \sim r$
3. If $p \sim q$ then $q \sim p$.
4. $\top \not\sim \bot$
5. $\top \sim p \rightarrow p$
6. $\bot \sim p \rightarrow \bot$
7. If $p \sim q$ then $\neg p \sim \neg q$.
8. If $\forall q(p \sim q \rightarrow p = q)$ and $q \sim r$ then $p \wedge q \sim p \wedge r$ and $p \vee q \sim p \vee r$.
9. If $p \wedge r = \bot$, and $q \wedge s = \bot$, $p \sim q$ and $r \sim s$ then $p \vee r \sim q \vee s$.

Of these I think 1-6 are pretty straightforward. 7-9 are perhaps best demonstrated via examples. For example 7 grounds the intuition that the proposition that 15 is not nice$_*$ and the proposition that 15 is not nice* are indiscernible. 8 and 9 are slightly harder to motivate. Intuitively the proposition that snow is white and 15 is nice* is indiscernible from the proposition that snow is white and 15 is nice$_*$ (and similarly for the disjunctive propositions.) This intuition is grounded by the general principle 8 assuming that the proposition that snow is white is totally determinate (given the vagueness of whiteness it presumably isn't a plausible assumption.) Our goal, then, is to give a formal account of $\sim$ which validates these intuitive principles.

One thing to notice is that the restriction that r is totally determinate is required in principle 8. Without the restriction 8. would be false: if we let r and p be the proposition that 15 is nice* and q be the proposition that 15 is nice$_*$ we get that the proposition that 15 is nice* and nice* is indiscernible from the proposition that 15 is nice* and nice$_*$. However the latter conjunction is inconsistent whereas the former is equivalent to the proposition that 15 is nice*. The problem here is that we have not switched the proposition that 15 is nice* for the proposition that 15 is nice$_*$ *uniformly*. This problem is the key to motivating the formal account to follow.

The central guiding principle, for calculating whether complex propositions are indiscernable from one another, is the principle that if p and q are indiscernible, then uniformly switching p for q in any other proposition results in an indiscernible proposition. The notion of uniform substitution and uniform switching is easy to make sense of if we are talking about sentences instead of propositions: for example if ϕ is the result of replacing the sentence '15 is nice*' (or, the term 'i') everywhere with the sentence '15 is nice$_*$' (or the term '$-i$') in ψ then ϕ is obtained from ψ by uniform substitution. And if you simultaneously

(not sequentially) substitute '15 is nice$_*$' for '15 is nice*' as well, then ϕ is obtained from ψ by uniformly switching '15 is nice$_*$' and '15 is nice*'. Consider:

1. 'snow is white or i is complex'
2. 'snow is white or $-i$ is complex'
3. '(snow is white and 15 is nice*) or 15 is nice$_*$'
4. '(snow is white and 15 is nice$_*$) or 15 is nice*'
5. '(snow is white and 15 is nice$_*$) or 15 is nice$_*$'

So 2. can be obtained from 1. both by uniformly substituting 'i is complex' for '$-i$ is complex' and uniformly switching them. 4., but not 5, can be obtained from 3. by uniformly switching '15 is nice*' for '15 is nice$_*$'. And 5., but not 4., can be obtained by a uniform substitution. It is only *switchings* that guarantee the sentences express indiscernible propositions – for example 1. and 2. express propositions that are indiscernible from one another, as do 3. and 4.; however 4. and 5. do not.

This is all very well and good, but it is not so clear, however, what a uniform switching of *propositions* within another proposition amounts to since we can't appeal to the typographical features a sentence has. In order to solve this problem we need to introduce the mathematical notion of an automorphism:

Definition 5.2.0.1. *Let $\mathbb{B}$ be a complete Boolean algebra (such as the algebra of propositions.) Then $\sigma : \mathbb{B} \to \mathbb{B}$ is an automorphism of $\mathbb{B}$ iff*

- *σ is a bijection (a one to one correlation.)*
- *$\sigma(\neg p) = \neg\sigma(p)$ for each $p \in \mathbb{B}$*
- *$\sigma(\bigvee X) = \bigvee\{\sigma(p) \mid p \in X\}$ for $X \subseteq \mathbb{B}$*

It follows from the definition that $\sigma(\bigwedge X) = \bigwedge\{\sigma(p) \mid p \in X\}$ for $X \subseteq \mathbb{B}$. An automorphism is therefore a bijective function on a Boolean algebra that preserves negation and arbitrary disjunctions and conjunctions.

Let p be the proposition that 15 is nice* and q the proposition that 15 is nice$_*$. Given some natural assumptions there will be an automorphism, σ, which switches p and q, i.e. $\sigma(p) = q$ and $\sigma(q) = p$, which also fixes every determinate proposition. Let r be the proposition that snow is white, so, idealising somewhat, r is a totally determinate proposition. We claimed that that 3. and 4. expressed indiscernible propositions, i.e. that $(r \wedge p) \vee q) \sim (r \wedge q) \vee p)$. In fact there is an automorphism that maps one to the other – and furthermore, it's the very automorphism we just considered: σ. $\sigma((r \wedge p) \vee q)) = \sigma(r \wedge p) \vee \sigma(q) = (\sigma(r) \wedge \sigma(p)) \vee \sigma(q) = (r \wedge q) \vee p)$.

There is an important distinction to bear in mind: some, but not all, automorphisms only map indiscernible propositions to indiscernible propositions. We can work under the assumption that there is a more restricted set of automorphisms, G, on the algebra of propositions which preserve further features of one's rational propositional attitudes. Call these automorphisms the 'admissible' automorphisms. Given an admissible automorphism, σ, p will be indiscernible from $\sigma(p)$. This means we can reverse the order of definitions, so that given a set of admissible automorphisms, G, we can say that p and q are indiscernible iff $\exists\sigma \in G\ \sigma(p) = q$. Thus:

> It is totally determinate whether p iff $\sigma(p) = p$ for all admissible automorphisms σ. (5.14)
>
> It is determinate that p iff, for each admissible σ, $\sigma(p)$. (5.15)

An equivalent and useful way to formulate this condition is to think of determinacy as an operator in the object language, Δ, whose interpretation is given by the function $\delta(p) :=$

$\bigwedge\{\sigma(p) \mid \sigma \in G\}$. There is a special name for the set $\{\sigma(p) \mid \sigma \in G\}$: it's called the *orbit* of p, which we can write $Orb(p)$. The interpretation of the determinacy operator is therefore given by $\bigwedge Orb(p)$.

5.2.3 Admissible automorphisms

Note that an automorphism of propositions will preserve all the inferential connections between those propositions, thus the logical connections between the beliefs (or more generally the Vings) that $p_1, ..., p_n$ are preserved under automorphisms. Thus automorphisms preserve one of the aspects of a proposition's attitudinal role that we are interested in. However, not all automorphisms preserve the relevant features - there is more to the notion of indiscernibility than having symmetrical logical roles. For example, under plausible assumptions there is an automorphism that maps the proposition that there are an even number of books on my bookshelf to its negation.[4] These two propositions, however, are not indiscernible since they play a very different epistemic role. For example they are clearly not evidentially symmetric: I might look at my bookshelf and see that there are only two books. It is clearly possible that I be in this evidential state, in which the evidential probability that there are an even number of books on my bookshelf is much greater than the probability that there are an odd number. The evidential relationship between the proposition that 15 is nice* and the proposition that 15 is nice$_*$ is relevantly disanalogous – I could never have evidence for one and not the other (indeed, in this case, I could never have evidence for either) – an adequate account of indiscernibility should take this into account.

The intuitive gloss is that two propositions are indiscernible iff one cannot coherently hold a certain kind of propositional attitude towards one proposition but not the other. One could put this slightly more contentiously as saying that p and q are indiscernible if a belief or desire that p has the same conceptual role as a belief or desire that q.

Our task is to say a bit more about what makes an automorphism admissible. I should begin by clarifying that I do not mean to say that one couldn't in principle believe, or hold some attitude, towards one but not the other of a pair of indiscernible propositions. I imagine that if one was careless enough one could even end up believing contradictions. What I mean to say is that no rational person could coherently hold distinct doxastic attitudes towards these propositions, just as one could not rationally believe a contradiction.

The characteristic epistemological properties of logical truths and falsehoods (understood as properties of propositions) are features that rational agents exhibit. Similarly the epistemological properties of indiscernible propositions should be explicated in terms of the belief forming dispositions of a perfectly rational agent. A natural place to start explaining what it means for one proposition to be 'interchangeable' for another would be to start with a theory of rational propositional attitudes. In the case of desire and belief there is already a very promising theory of rational preferences due to Jeffrey and Bolker which we outlined in chapter 3.

Now to turn to the matter at hand. Instantiating preferences which comply with the principles of Jeffrey-Bolker decision theory is necessary, but not sufficient, for being coherent. No-one could coherently care whether 15 is nice* or not, nor could they possibly have evidence which would rationally permit adopting a higher credence in the proposition that 15 is nice* than in the proposition that 15 is nice$_*$, even though they may comply with the Jeffrey-Bolker axioms. Phrasing these constraints in terms of preferences suggests a promising definition of admissibility:

(Ad) σ is admissible (if and) only if, for any coherent preference ranking $\preceq$ and proposition p, p and $\sigma(p)$ are ranked together: $p \approx \sigma(p)$.

[4]The assumption is that this proposition is a disjunction of atoms, and that there are just as many atoms beneath this proposition as not. The former assumption is universally made among possible world semanticists, so the second assumption reduces to their being just as many worlds in which there is an even number of books on my bookshelf as not.

Unfortunately this definition will not quite do if we want this account to remain compatible with the view defended in chapter 2. It is worth reflecting on why this is. Let S, N^* and N_* be the propositions that snow is white, 15 is nice* and nice$_*$ respectively. We have argued that $S \vee N^*$ and $S \vee N_*$ are indiscernible on the simplify assumption that S is totally determinate. However, $Pr(S \vee N^* \mid S \vee N_*) < 1$ if $Pr(S) < 1$ and S and N^* are independent, yet clearly $Pr(S \vee N_* \mid S \vee N_*) = 1$. If $S \vee N_*$ could be someone's total evidence these indiscernible propositions would be discernible to this person given their evidence; if this could be someone's evidence then these indiscernible propositions will feature differently in preferences of an agent who has this evidence. Now the question is, could this conditional probability be a rational person's credence? I.e. could $S \vee N_*$ be the strongest evidence a person has at a given time (as opposed to the stronger S say)?

With this particular example it does not seems very likely. However in chapter 2 I argued that when our evidence is inexact the strongest evidence we have is a *vague* proposition. Since vagueness is a type of indeterminacy it seems that in general we can be in a situation where our strongest evidence is a determinate but not totally determinate proposition. According to the claims of chapter 2 (which, note, are not forced on us by anything we say in this chapter) contingently indeterminate propositions (in this case 'mixtures' of pure indeterminate propositions and totally determinate propositions) are not always indiscernible amongst rational agents who have learnt something.[5]

In order to allow for inexact evidence to be vague evidence we must restrict attention to *initial* preference rankings – rankings of a conceptually coherent agent who has not yet received any evidence.[6]

(Ad) σ is admissible (if and) only if, for any coherent *initial* preference ranking $\preceq$ and proposition p, p and $\sigma(p)$ are ranked together: $p \approx \sigma(p)$.

As a consequence of this note that if σ is admissible and $\preceq$ is a coherent initial preference ranking we have:

$p \preceq q$ iff $p \preceq \sigma(q)$.

$p \preceq q$ iff $\sigma(p) \preceq q$.

$p \preceq q$ iff $\sigma(p) \preceq \sigma(q)$.

Thus a rational persons initial preferences are completely invariant under admissible automorphisms. Either of the first two principles entails (Ad) and could serve as an alternative definition of an admissible automorphism.

Someone might object to my definition on the basis that my primitives are mysterious – how do these constraints on coherent *priors* and initial preferences manifest themselves for informed people? In this respect, the definition has some fairly substantive consequences. Say that $\prec'$ is a coherent preference ranking if and only if for some proposition, e, such that it is possible that e is someone's total evidence, and some coherent initial preference ranking $\prec$, $\prec' = \prec_e$. Finally, given a relation on propositions, R, define $\sigma(R) := \{\langle \sigma(p), \sigma(q)\rangle \mid \langle p, q\rangle \in R\}$. We then get the following fact:

If $\prec$ is a coherent preference ranking, $\sigma(\prec)$ is also a coherent preference ranking.

In order to show this fact we assume that if it is conceptually possible that e is someone's total evidence and e' is an indiscernible proposition, $e \sim e'$, then it is conceptually possible that e' is someone's total evidence. Thus, for example, it's possible that Harry is bald is

[5]Say that a proposition p is a *pure* indeterminate proposition if the conjunction of all propositions indiscernible from p is the contradictory proposition. It seems that pure indeterminate propositions, such as the proposition that 15 is nice* could not be someone's total evidence.

[6]This fact is just a restatement of the issue we encountered in chapter 2: we had to be careful to formulate the supervenience of credences in the vague in terms of credences in the precise in terms of *coherent priors*, i.e. regular probability functions that are conceptually coherent.

someone's total evidence just in case it's possible that Harry is bald* is someone's total evidence, where the proposition that Harry is bald* is some proposition indiscernible from the proposition that Harry is bald. This assumption about evidence would be entailed by the claim that, if e is indiscernible from e', the proposition that my total evidence is e is indiscernible from the proposition that my total evidence is e'. Neither of these assumptions is entailed by (Ad) alone. If our assumption is not true then we should amend our definition of an admissible automorphism: an automorphism is admissible just in case it both preserves all coherent initial preference rankings and maps possible evidence to possible evidence.

It should be stressed that the operative notion of possibility here is not metaphysical possibility: if $e \sim e'$ and it's metaphysically possible that my total evidence is e it does not follow that it's metaphysically possible that my total evidence is e'. Suppose that e is the conjunction of some precise evidence and the proposition that Harry is bald, and e' is the conjunction of the same precise evidence with the proposition that Harry is bald*. According to chapter 2 it's metaphysically possible that my total evidence is e or at least something indiscernible from e. Yet presumably the fact that my total evidence is e supervenes on facts that would be present if my total evidence were to be e'; the way I obtain inexact evidence that Harry is bald* is exactly the same as the way in which I obtain inexact evidence that Harry is bald (through imperfect vision, say.) In a situation in which the agents total evidence is p and the proposition that Harry is bald it's also epistemically possible that the agents total evidence is p and the proposition that Harry is bald* – however this is just the kind of epistemic possibility that is open to us when vagueness is involved. It's vague which of a number of indiscernible propositions is the agents evidence, although it's determinate that whatever it is, it's a determinately true, but not totally determinate, proposition.

We may now turn to a consequence of having invariant initial preferences. Suppose that $\prec'$ is a coherent preference ranking so $\prec'=\prec_e$ for some possible evidence e and initial coherent ranking $\prec$. By definition $p \prec_e q$ iff $p \wedge e \prec q \wedge e$. So for any admissible automorphism, σ, $p \prec' q$ iff $p \wedge e \prec q \wedge e$ iff $\sigma(p \wedge e) \prec \sigma(q \wedge e)$, since $\prec$ is initial, iff $\sigma(p) \wedge \sigma(e) \prec \sigma(q) \wedge \sigma(e)$ iff $\sigma(p) \prec_{\sigma(e)} \sigma(q)$. Thus $\sigma(\prec') = \prec_{\sigma(e)}$ so $\sigma(\prec')$ is a coherent ranking too.

Let us move on to other consequences of our definition. It is easy to see that the identity mapping is an automorphism and satisfies (Ad). In the appendix we show also that compositions and inverses of admissible automorphisms are admissible, so G must be a group. The *orbit* of a proposition can defined $Orb(p) := \{\sigma(p) \mid \sigma \in G\}$. It is easy to check that $\bigvee Orb(p)$ and $\bigwedge Orb(p) (= \delta(p))$ are fixed by every automorphism in G, so both these propositions are totally determinate (see appendix.)

Note also that (Ad) entails that for every coherent prior $Pr(p) = Pr(\sigma(p))$ (see appendix for argument.) So the following account explains why no-one can rationally be certain in the proposition that 15 is nice*.

All in all (Ad) gives us a very strong constraint on which automorphisms are admissible. Perhaps this constraint is maximally strong in that it extensionally characterises the notion of indiscernibility.

One might wonder whether the involvement of desires, through preferences, is really necessary for our characterisation of indiscernibility. It is worth considering whether one could have just considered automorphisms which preserve rational belief and other doxastic states. Would it have been enough, for example, to stipulate that an admissible automorphism, σ, must only satisfy the equation $Pr(p) = Pr(\sigma(p))$ for every coherent prior probability function? A potential source of counterexamples to this claim comes from the existence of unverifiable propositions. There does not seem to be any possible evidence for the proposition that there is an invisible undetectable teapot on the north pole and the proposition that there is an invisible teapot on the south pole; yet these are not indiscernible propositions.

This objection partly rests on a misunderstanding of the notion of a coherent prior I tried to explicate in chapter 1. While it may, in some sense, be unreasonable to have priors

which assign a high probability to the proposition that there is an invisible undetectable teapot on the north pole, there is nothing conceptually *incoherent* about this belief. It is not, for example, like having priors that assign the proposition that there's a there is a male vixen in the north pole a high probability.

Even if one does not accept the weak notion of rationality I've been calling 'coherence' once bouletic attitudes are included the picture changes further. I see no reason why one could not coherently *care* whether there is an invisible teapot at the south pole, but it seems that the same could not be said about caring whether 15 is nice*. This response requires a moderate kind of permissivism about rational desire, but I do not think that is too objectionable. While there is obviously a clear sense in which there is something wrong with this kind on desire, it is not *incoherent* in the sense in which the desire to be a married bachelor is, for example. My thesis is that it is incoherent to care intrinsically about whether 15 is nice* or not in exactly the same way as it is incoherent to want to be a married bachelor.

I believe the notion of coherence that I am appealing to has some pretheoretic appeal. Its relation and importance to the study of vagueness has already been discussed in literature. For example, in [41], Hartry Field describes the example of Roger, who thinks that if his bank account password has the same last digit as the number of nanoseconds Bertrand Russell was old for, then his life will go better, and Sam, who things that his life will go better if the last digit of his bank account password is the seventeenth significant digit of the Centigrade temperature at the currently hottest point in the interior of the sun. According to Field, while Sam's belief is thoroughly irrational, Roger's is intuitively even worse as it is conceptually confused.

An admissible automorphism of propositions preserves the coherence of all beliefs and desires in those propositions. There is a natural question as to whether there are further propositional attitudes whose coherence is preserved by these automorphisms. For example, in the same paper, Field suggests several propositional attitudes that one could not coherently hold towards an indeterminate proposition, which includes, as well as bouletic and doxastic attitudes, hoping, wondering and moral attitudes [41]. For a critical discussion of this claim see also [135].

Although the term 'conceptual role' comes with a lot of baggage which I don't want spend time on, it is certainly a term one can make sense of without the baggage, and one that does a good job of describing what I have in mind here. I take it that the conceptual role of a belief or desire that p includes at least the role that belief or desire would play in a somewhat idealised psychological description of how a coherent rational agent's beliefs and desires would evolve over time in response to sensory inputs, and the role these desires and beliefs in turn affect their dispositions to act in certain ways. It seems very plausible, given our preference theoretic definition of admissibility, that a belief or desire that p has the same conceptual role as a belief or desire that $\sigma(p)$, for admissible σ. Although not as direct, the notion of conceptual role presumably extends to other propositional attitudes such as hopes, fears, wonderings and so on. Therefore a natural further question of interest is whether admissible automorphisms preserve the conceptual role of these attitudes. For example, letting σ be an admissible automorphism, whether the following holds:

> A hope, fear, wondering that/whether p has the same conceptual role as a hope, fear, wondering that/whether $\sigma(p)$. (5.16)

These additional theses bears further investigation; I shall not do so here.

Although this account of indeterminacy is not linguistic it is in agreement with many of its consequences. Since indiscernible propositions play functionally symmetric roles in the network of one's rational beliefs and desires, any non-trivial permutation of those beliefs and desires would not make a difference at the level of rational decision making. Insofar as one is behaving rationally, one's preferences and desires between indiscernible propositions is inconsequential; redistributing one's beliefs and desires among indiscernible propositions

does not affect one's linguistic and non-linguistic behaviour. Insofar as the purpose of communication is to bring about certain beliefs, desires and behaviour in our audience by affecting their beliefs, the truth values of indeterminate propositions cannot be relevant for rational communication. If s expresses an indeterminate proposition, p, it is not permissible to assertively utter s knowing that p is indeterminate. Since s is equi-assertible with any sentence expressing a proposition indiscernible from p, then one can know that s is one of a class of equi-assertible sentences, one of which is false.

5.2.4 Explanatory priority

Let us now turn to the question of whether this account is really an informative theory of indeterminacy. I should start off by saying that it is not informative purely in virtue of its having reduced the concept of indeterminacy to the concept of a coherent prior credence and initial preference. Someone might rightly object that one needs to have *some* grasp of the notion of indeterminacy to understand why it's incoherent to believe that Harry is bald when he has some borderline number of hairs. This much it shares with many philosophical theories. For example David Lewis's theory of counterfactuals made use of the concept of similarity between worlds. Yet the concept of similarity employed in this analysis is hard to explain to someone who does not already have the ability to evaluate counterfactuals; Lewis's theory of counterfactuals cannot really be said to provide a reduction of counterfactual to non-counterfactual facts, even if it is a reduction of counterfactual to similarity facts.

It's worth stressing a point we made earlier. We don't really *need* this very weak notion of rationality – having a coherent credence – for an adequate account of indeterminacy provided we are reasonably permissive about what you can care about. This point noted, it nonetheless seems as if the concept of coherence has some independent standing, just as we have some kind of independent grasp of the concept of counterfactual similarity. The notion of coherence was at least apparent to philosophers who drew a distinction between the epistemic status of conceptual truths and merely necessary and other non-conceptual truths.[7]

The theory is informative in virtue of the connections it draws with other concepts we already understand well enough: evidence, rational preferences, beliefs, knowledge, and so forth. Unlike a very primitivist theory of indeterminacy (such as Barnett [4], for example), this account situates the concept of indeterminacy within a large circle of related concepts. Unlike the anti-reductionist view defended by Field in [37], say, we can also turn these implicit connections into an explicit definition in terms of these other concepts. A truly reductionist view, one that does not appeal to any notion which can be explained to someone who does not already have the concept of indeterminacy, I think, is not to be found. The same could be said about countless other concepts of interest to philosophers (e.g. counterfactuals, causation, knowledge, et cetera.)

However, there is an objection along these lines that seems to be more specific to this account. 'Surely', the imagined objection goes, 'the true analysis of indeterminacy is supposed to *explain* the normative rules governing one's propositional attitudes towards indeterminate propositions – any analysis that must rely on taking such rules as basic has made a wrong turn somewhere.' Let me record that I have my doubts about the claims of explanatory priority, especially when semi-technical terms like 'it's indeterminate whether' are concerned. I do not see the project I am engaging in so much as one of conceptual analysis – subject to linguistic data (presumably to be taken primarily from philosophy journals) – but as a search for a notion that plays the theoretical and explanatory role that talk of indeterminacy is put to in philosophy.

With that said, there is some force to the objection – clearly *some* 'shoulds' require

[7] Although, it should be stressed, I am not analysing the epistemic distinction the way it is traditionally analysed: in terms of one's competence with words or concepts.

explanation: for example, you shouldn't cross the road without looking because you're likely to get hit by a car. But explanations like these rely more primitive fact about 'shoulds'; maybe you should care about not being in pain, or you should do things that bring about things you care about. The question 'why should I bring about things I care about' seem to be misguided; it sounds as if the questioner is asking for an explanation of an ought from an is. 'Why should I care about maximising expected utility' seems also misguided: utility is just whatever it is that you should care about; but it's hard to say anything general that sounds like an explanation of why we should maximise expected utility.

I think that the question 'why is it incoherent to believe that not all vixens are female foxes?', or the question 'why is it incoherent to care intrinsically about whether Harry is bald?' are similarly misguided questions on the current account of propositions. According to this account, propositions are *individuated* by the coherence of beliefs and desires concerning them. It is rather like asking a possible worlds theorist 'why is the proposition that Hesperus is bright true in this world and not that world?', or 'why are there no consistent propositions entailing the proposition that Hesperus is not Phosphorus?'; propositions are simply individuated by facts like these. A better question would be 'Why does the sentence 'Hesperus is Phosphorus' express this proposition rather than that one?' Similarly, a better question in this context would be 'why does the sentence '15 is nice*' express a proposition with this attitudinal role, and not this one?' In which case, we could say something informative about the way in which the predicate 'nice*' was introduced. Since, for any possible attitudinal profile, there is some proposition with that attitudinal profile, the latter kind of question makes more sense than the former.

5.3 The framework

Note that the identity automorphism trivially satisfies the condition (Ad). Furthermore, if σ and τ are automorphisms on the algebra of propositions that preserve all coherent initial preferences, then $\sigma \circ \tau$, the composition of σ and τ, and σ^{-1}, the inverse of σ will also do so. Thus if Σ is the set of all admissible automorphisms Σ will be a group.

Most of the applications mentioned in §5.1, however, require more specific notions than this. While our determinacy operator gives us a good analysis of claims like (5.17),

If it's vague whether p then there's no fact of the matter whether p (5.17)

which is a statement which can be used to distinguish a classical account of vagueness from epistemicism, the converse does not hold which means we cannot generate an analysis of vagueness on this basis alone. There is plausibly no fact of the matter whether the truth-teller is true or not, however this claim is not *vague* since it lacks a corresponding Sorites sequence. These are all *species* of indeterminacy, and each species will have to be analysed separately.[8]

However, each species of indeterminacy can be formulated in our semantic framework by placing some restriction on the set Σ. Thus, for example, to model vagueness one would want to restrict one's attention to elements of Σ which of conceptual necessity fix the truth value of every precise proposition.[9] Everything fixed by every element of Σ is precise, but some precise propositions are not fixed by every element of Σ such as, perhaps, the continuum hypothesis, or certain counterfactual statements.

[8]It is arguable in each case whether the putative species is in fact a kind of indeterminacy. The Principal Principle, if true, provides one way in which propositions about the future can be distinguished from each other evidentially, which is something defenders of indeterminacy about the open future would need to explain on my account.

[9]Given any Kripke frame, $\langle W, R\rangle$, for an operator representing a species of indeterminacy, we can generate a corresponding set of automorphisms as the elements σ of Σ such that for every (im)possible world $w \in W$, and $p \subseteq W$, if $R(w) \subseteq p$ then $w \in \sigma(p)$. We shall show in the appendix that not every Kripke frame is structurally suitable to model a species of indeterminacy - there are Kripke frames which generate a set of automorphisms counting more things as determinate than the frame.

Clearly this restriction won't help us determine what the precise propositions *are*, but it is a substantive claim about vagueness nonetheless. For example, it ensures certain facts about the logic of vagueness and ensures that claims like (5.17) come out true. It is substantive in another sense too: not every operator characterised by an accessibility relation can be characterised by some subset of Σ (see failures of the converse of Lemma 5.3.3 in section 2.)

In this section a number of technical questions concerning the account of indeterminacy will be addressed. The most important technical concept is that of an automorphism:

Definition 5.3.0.2. *Let $\mathbb{B}$ be a complete Boolean algebra. An automorphism of $\mathbb{B}$ is a bijective function, $\sigma : \mathbb{B} \to \mathbb{B}$, such that:*

1. $\sigma(\neg b) = \neg\sigma(b)$
2. $\sigma(\bigsqcap_{b\in X} b) = \bigsqcap_{b\in X} \sigma(b)$

Here $\sqsubseteq$ represents the Boolean ordering and $\bigsqcap, \bigsqcup$ represent the infimum and supremum operations respectively . We begin by stating without proof some well known facts about automorphisms.

Proposition 5.3.1. *If σ and τ are automorphisms of $\mathbb{B}$:*

1. *σ^{-1} and $\sigma \circ \tau$ are automorphisms.*
2. *$a \sqsubseteq b$ iff $\sigma(a) \sqsubseteq \sigma(b)$ for all $a, b \in \mathbb{B}$.*

Proposition 5.3.2. *If $\mathbb{B} = \mathcal{P}(W)$ for some set W then*

1. *For any permutation t of W (i.e. any bijection $t : W \to W$), the function $\sigma(X) \mapsto \{t(x) \mid x \in X\}$ is an automorphism.*
2. *For any automorphism σ of $\mathbb{B}$, there is some permutation, t, on W such that $\sigma(X) = \{t(x) \mid x \in X\}$ for all $X \subseteq W$, given by setting $t(x) = y$ iff $\sigma(\{x\}) = \{y\}$.*
3. *Composing and inverting permutations corresponds to composing and inverting their corresponding automorphisms, and vice versa.*

I shall use σ interchangeably to denote an automorphism or its corresponding permutation on W if no ambiguity is possible. From here on I shall mostly limit discussion to powerset algebras, so that a closer comparison with modal logic can be drawn. I shall occasionally note how things can be generalised to arbitrary complete Boolean algebras when applicable.

Definition 5.3.2.1. *An indeterminacy species is a pair $\langle W, \Sigma\rangle$ where Σ is a non-empty set of automorphisms of $\mathcal{P}(W)$. A model is a triple $\langle W, \Sigma, \Vdash\rangle$, where $\Vdash$ is a relation between W and a monomodal propositional language such that;*

- *$w \Vdash \phi \wedge \psi$ iff $w \Vdash \phi$ and $w \Vdash \psi$*
- *$w \Vdash \neg\phi$ iff $w \nVdash \phi$*
- *$w \Vdash \Box\phi$ iff $w \in \sigma(\{x \mid x \Vdash \phi\})$ for all $\sigma \in \Sigma$.*

A sentence is valid in a model $\langle W, \Sigma, \Vdash\rangle$ just in case $w \Vdash \phi$ for all $w \in W$, and valid over a species $\langle W, \Sigma\rangle$ just in case it is valid in every model of the form $\langle W, \Sigma, \Vdash\rangle$.

A general indeterminacy species is a pair $\langle \mathbb{B}, \Sigma\rangle$, where $\mathbb{B}$ is an arbitrary complete Boolean algebra, Σ is a non-empty set of automorphisms of $\mathbb{B}$. A model over a general species is a triple $\langle \mathbb{B}, \Sigma, [\![\cdot]\!]\rangle$ where $[\![\cdot]\!]$ is a function between modal proposition language and $\mathbb{B}$ such that

Name	Principles	Sufficient species condition
D	$\Box p \to \Diamond p$	$\Sigma \neq \emptyset$
T	$\Box p \to p$	$id \in \Sigma$
4	$\Box p \to \Box\Box p$	$\sigma, \tau \in \Sigma \Rightarrow \sigma \circ \tau \in \Sigma$
5	$\Diamond p \to \Box\Diamond p$	$\sigma, \tau \in \Sigma \Rightarrow \sigma^{-1} \circ \tau \in \Sigma$
B	$p \to \Box\Diamond p$	$\sigma \in \Sigma \Rightarrow \sigma^{-1} \in \Sigma$
B^n	$p \to \Box\Diamond^n p$	$\sigma \in \Sigma \Rightarrow \exists \tau_1, \ldots \tau_n \in \Sigma, \sigma^{-1} = \tau_1 \circ \ldots \circ \tau_n$
G1	$\Diamond\Box p \to \Box\Diamond p$	$\sigma, \tau \in \Sigma \Rightarrow \sigma \circ \tau = \tau \circ \sigma$

Figure 5.1: Some common principles and species conditions which suffice for their validity

- $[\![\phi \wedge \psi]\!] = [\![\phi]\!] \sqcap [\![\psi]\!]$
- $[\![\neg\phi]\!] = \neg[\![\phi]\!]$
- $[\![\Box\phi]\!] = \bigsqcap_{\sigma \in \Sigma} \sigma([\![\phi]\!])$

I shall call members of W worlds. We shall introduce a defined operator $\Diamond p := \neg\Box\neg p$. It is easy to show that $w \Vdash \Diamond\phi$ iff $w \in \sigma(\{x \mid x \Vdash \phi\})$ for some $\sigma \in \Sigma$, since for any automorphism $\sigma(W - X) = W - \sigma(X)$.

A few words are in order concerning the relation between the formal machinery and the informal account of §5.2. Every informal species of indeterminacy, S, whether that be vagueness, the open future, or what have you, is associated with a restricted set of conceptual-role preserving automorphisms Σ which is used to determine the S-determinate facts at every (im)possible world of the model. This rules out the possibility that the restriction on a species is infected with that very species of indeterminacy. This assumption is justified when the restriction in place is an S-determinate one: if S is our species of indeterminacy, the restriction is that, of epistemic/logical necessity, every element of Σ fixes the truth value of every S-determinate proposition. The notion of epistemic or logical necessity here is born out formally in terms of truth at every (im)possible world of the model. However, while this might be a safe assumption for some species of indeterminacy, there are reasons relating to the paradoxes of higher order vagueness to think that even logical and epistemic necessity are vague. To account for this possibility one would need to have a larger class of points (including logically impossible points) which each have a different set Σ associated with them. For our purposes it can be noted that higher order vagueness can still be modelled in this framework (since the modalities can iterate in non-trivial ways) and that the more general framework would subsume this formalism as a special case. Although a more general notion of frame, in which each world has its own set of automorphisms, is desirable although I have not yet worked out the details.

Our next task is to find a suitable logic for indeterminacy species. By way of analogy, we want a logic which stands to species as the logic K stands to Kripke frames. It is easy to check that the modal logic KD (see figure 1) is valid on every species. In the next few lemmas I shall argue that in fact KD is exactly the logic of species. Other logics can be obtained by putting further constraints on species. See figure 1 for conditions which guarantee some familiar logics.

Notice that D, $\Box p \to \Diamond p$, is valid in every species due to the fact that we stipulated that Σ was non-empty. It is natural to wonder what would happen if we relaxed that constraint. Would we, for example, get a logic of exactly K instead? The logic of the species with an empty set of automorphisms is very easy to axiomatise, and consists of adding to K the axiom $\Box p$.

This marks a difference between Kripke semantics and species semantics. The formula $\Box(\Box p \to \Diamond p)$ is valid in every species, even if we allow Σ to be empty: if Σ is empty, everything is necessary, and if it's not then D is valid, and validity is preserved under

necessitation. However $\Box(\Box p \to \Diamond p)$ is not provable from K alone since it is refutable in any frame with a point that sees a dead end.[10] Thus the logic of empty and non-empty species is strictly between K and KD. So in general there are normal modal logics which are characterisable by some class of Kripke frames but no class of species.

Now to show that the logic of indeterminacy species is KD. Since we do not have any purpose-made way of giving completeness proofs for logics of indeterminacy, I shall begin by relating species to frames.

Lemma 5.3.3. *Every indeterminacy model, $\langle W, \Sigma, \Vdash_1 \rangle$, determines a unique Kripke model, $\langle W, R, \Vdash_2 \rangle$, based on the same set of worlds, such that $\Vdash_1 = \Vdash_2$.*

Proof. We define Rxy iff $\sigma(y) = x$ for some $\sigma \in \Sigma$, and for atomic sentences let $x \Vdash_2 p$ iff $x \Vdash_1 p$.

We then show by induction that truth at a world coincides in both models. The crucial clause is for $\Box$.

Let $[\![\phi]\!]_i := \{x \mid x \Vdash_i \phi\}$. Assume, for induction, that $[\![\phi]\!]_1 = [\![\phi]\!]_2$.

Suppose that $x \in [\![\Box\phi]\!]_1$. So $x \in \sigma([\![\phi]\!]_1)$ for all $\sigma \in \Sigma$. Suppose that Rxy, i.e. that $\exists \tau \in \Sigma$ with $\tau(y) = x$. Since $x \in \tau([\![\phi]\!]_1)$, $\tau^{-1}(x) = y \in [\![\phi]\!]_1 = [\![\phi]\!]_2$. So $x \in [\![\Box\phi]\!]_2$.

Now suppose that $x \in [\![\Box\phi]\!]_2$, so $y \in [\![\phi]\!]_2$ for all y such that Rxy. Let $\sigma \in \Sigma$. There is some y such that $\sigma(y) = x$ since σ is (corresponds to) a permutation on W. Also $y \in [\![\phi]\!]_2$, since Rxy, so $y \in [\![\phi]\!]_1$ by the inductive hypothesis. Thus $x \in \sigma(\{y\}) \cup \sigma([\![\phi]\!]_1) = \sigma(\{y\} \cup [\![\phi]\!]_1) = \sigma([\![\phi]\!]_1)$ as required for $x \in [\![\Box\phi]\!]_1$. $\Box$

It does not seem to be possible to generalise this result to general indeterminacy species. In this case the general indeterminacy species would determine a general Kripke frame (see, e.g., Blackburn et al [10] p28), i.e. a Kripke frame equipped with a certain set of subsets of W which serve as the possible interpretations of the propositional letters (that set will be determined by $\mathbb{B}$'s representation as an algebra of sets via Stone duality.)

It should be noted that the converse of lemma 5.3.3 does not hold: not every Kripke frame corresponds to a species. For example there are only two permutations on the set $\{0, 1\}$, the identity permutation and the one that swaps 0 and 1. There are thus only three species based on that set corresponding to the Kripke frames $\langle \{0,1\}, \{0,1\}^2 \rangle$, $\langle \{0,1\}, \{\langle 0,1\rangle, \langle 1,0\rangle\} \rangle$ and $\langle \{0,1\}, = \rangle$. Thus no indeterminacy species corresponds to the Kripke frame $\langle \{0,1\}, \{\langle 0,1\rangle\} \rangle$. This mismatch can make a difference to the logics characterised by frames and species - some logics that can be characterised by classes of Kripke frames cannot be characterised by any class of species, indicating that such logics do not correspond to possible species of indeterminacy. We have seen this already when we considered allowing the empty species.

Theorem 5.3.4. *A sentence is a theorem of KD if and only if it is valid in every species.*

Proof. Suppose that Γ is KD consistent. Then by standard methods (see [10] proposition 2.15) every element of Γ is true at the root of a tree-like Kripke model, $\langle W', R', \Vdash \rangle$, obtained by unravelling the canonical model. Since the canonical model for KD is serial, every branch of this tree will be infinite. Let / be the root of the tree.

Let $W = W' \cup \mathbb{N}$ and let Rxy iff $R'xy$ or $(x = 0$ and $y = a)$ or $(x = n + 1$ and $y = n)$. For $n \in \mathbb{N}$ assign truth values to propositional letters arbitrarily: this does not affect the truth values of formulae at /, so that $a \Vdash \gamma$ for each $\gamma \in \Gamma$, even with the extra points added. Now notice that every maximal branch in this new tree-like structure is infinite in both ways, i.e., each branch is isomorphic to $\mathbb{Z}$. For each branch b in $\langle W, R \rangle$, define an automorphism σ_b by the permutation that fixes every point of W not on the branch, and maps every point on the branch to the point directly beneath it, so if Rxy, and x and y are on b, $\sigma_b(y) = x$. We let our species be given by W and $\Sigma = \{\sigma_b \mid b$ a maximal branch over

[10] I do not know what the logic of species is if we include the empty species, but it would certainly be stronger than $\mathsf{K}+\Box(\Box p \to \Diamond p)$ since it would include $(\Box\bot \vee \Diamond^n \Box\bot) \to \Box\bot$ for each n.

$\langle W, R\rangle\}$. It is easy to check that the Kripke frame $\langle W, \Sigma\rangle$ determines, according to lemma 5.3.3, is $\langle W, R\rangle$. So by lemma 5.3.3, Γ is satisfiable on a species as required. □

Similarly we can show:

Theorem 5.3.5. *A sentence is a theorem of* KT *if and only if it is valid in every species containing the identity automorphism.*

The method very similar to theorem 5.3.4 so I shan't reproduce it.

Theorem 5.3.6. *A sentence is a theorem of* S5 *if and only if it is valid in every species* $\langle W, \Sigma\rangle$ *in which* Σ *is a group.*

Proof. Σ is a group iff $id \in \Sigma$, $\sigma^{-1} \in \Sigma$ whenever $\sigma \in \Sigma$ and $\sigma \circ \tau \in \Sigma$ whenever $\sigma, \tau \in \Sigma$. From figure 1 we can see that any species that satisfies these conditions validates KTB4, which is well known to be equivalent to S5.

For completeness we note that every species $\langle W, R\rangle$ where R is an equivalence relation corresponds to the species $\langle W, \Sigma\rangle$ where Σ is generated by the following set of permutations: $S = \{\sigma \mid \text{for any } x, y \in W, \text{ if } \sigma(x) = y,\ Rxy\}$. By proposition 5.3.2, it is sufficient to show that S contains the identity permutation and is closed under inverses, and function composition. $id \in S$ since R is reflexive. Suppose that $\sigma, \tau \in S$ and $\sigma \circ \tau(x) = y$. It follows that $R\tau(x)y$, and $Rx\tau(x)$ by the condition for membership in S, so by transitivity of R, Rxy. So $\sigma \circ \tau \in \Sigma$. Suppose that $\sigma \in \Sigma$, and that $\sigma^{-1}(x) = y$. It follows that $\sigma(y) = x$, so Ryx, and since R is symmetric, Rxy. Thus $\sigma^{-1} \in \Sigma$ as required.

Finally we note that S5 is complete with respect to Kripke frames based on equivalence relations. □

Let me end by mentioning an unsettled question: are KD4 and KD5 complete with respect to the properties given in figure 1?

5.4 Decision theory

Here we outline some basic facts about the set of admissible automorphisms and the constraints they place on initial preferences and priors. Given a relation, $\preceq$, write $p \prec q$ to mean that $p \preceq q$ and $q \not\preceq p$, and $p \approx q$ to mean that $p \preceq q$ and $q \preceq p$. The core of Jeffrey-Bolker decision theory can then be presented by the following maxims governing rational preference over a complete Boolean algebra with the bottom element removed, $\mathbb{B} \setminus \{\bot\}$:

1. $\preceq$ is a transitive total relation.
2. If $p \wedge q = \bot$ then
   (a) If $p \prec q$ then $p \prec (p \vee q) \prec q$
   (b) If $p \approx q$ then $p \approx (p \vee q) \approx q$
3. If $p \wedge q = \bot$ and $p \approx q$, then if $(p \vee r) \approx (q \vee r)$ for some r with $r \wedge p = r \wedge q = \bot$ and $p \not\approx r$, then $(p \vee r) \approx (q \vee r)$ for every such r.
4. If $p \prec R \prec q$ and $R = \bigvee_\alpha r_\alpha$ then for some β, $p \prec R_\beta \prec q$ where $R_\beta = \bigvee_{\alpha > \beta} r_\alpha$. Similarly if R is the infimum $\bigwedge_\alpha r_\alpha$.

given a proposition p and real valued random variable u we write the conditional expectation of u on p as $E(u \mid p)$.

Theorem 5.4.1. *If* $\prec$ *satisfies the above principles and* $\mathbb{B}$ *is atomless then there is a probability measure on* $\mathbb{B}$ *and a real valued random variable* u *on* $\mathbb{B}$ *such that* $E(u \mid p) < E(u \mid q)$ *iff* $p \prec q$.

Bolker's theorem shows that preferences can be represented by comparisons of the conditional expected utility.[11] In what follows I shall write $\preceq$ to denote a relation between propositions satisfying the Jeffrey-Bolker axioms above.

We begin by showing that the alternative definition of admissible is equivalent to that given by (Ad).

Definition 5.4.1.1. *Given a* set *of preference relations* Γ*, we say that the set of* ***admissible automorphisms*** *for* Γ*,* $ad(\Gamma)$ *is the set of automorphisms,* σ*, such that for any* $\preceq \in \Gamma$ *and any* p *and* q*:*

- $p \preceq q$ *if and only if* $p \preceq \sigma(q)$
- $p \prec q$ *if and only if* $p \prec \sigma(q)$

Proposition 5.4.2.

1. *For every* p*,* q *and* $\sigma \in ad(\Gamma)$*:* $p \preceq q$ *if and only if* $\sigma(p) \preceq q$*. Similarly for* $\prec$*.*
2. *For every* p*,* q *and* $\sigma \in ad(\Gamma)$*:* $p \preceq q$ *if and only if* $\sigma(p) \preceq \sigma(q)$*. Similarly for* $\prec$*.*
3. *For every* p *and* $\sigma \in ad(\Gamma)$*:* $p \approx \sigma(p)$*.*

Proof. 1. $p \not\preceq q$ iff $q \prec p$ (by linearity of $\preceq$). This happens iff $q \prec \sigma(p)$ since σ is admissible, and thus iff $\sigma(p) \not\preceq q$. The case for $\prec$ is similar.

2. $p \preceq q$ iff $p \preceq \sigma(q)$ (σ is admissible) iff $\sigma(p) \preceq \sigma(q)$ (by 1.)

3. $p \preceq p$ therefore $p \preceq \sigma(p)$ (admissibility of σ) and $\sigma(p) \preceq p$ (by 1.) therefore $p \approx \sigma(p)$ □

Proposition 5.4.3. $ad(\Gamma)$ *is a group, i.e.*

1. $id \in ad(\Gamma)$
2. $\sigma^{-1} \in ad(\Gamma)$ *if* $\sigma \in ad(\Gamma)$
3. $\sigma \circ \tau \in ad(\Gamma)$ *if* $\sigma, \tau \in ad(\Gamma)$

Proof. 1 and 3 are trivial given proposition 5.4.2.

For 2 suppose that $\sigma \in ad(\Gamma)$, and that $p \preceq q$. Let $q' = \sigma^{-1}(q)$. Then the following are equivalent:

$p \preceq q$

$p \preceq \sigma(q')$ (definition of q')

$p \preceq q'$ (σ is admissible)

$p \preceq \sigma^{-1}(q)$ (definition of q')

thus σ^{-1} is admissible. □

Throughout the rest of the section we shall assume that Γ denotes the set of relations that could be the initial preferences of a perfectly rational agent.

Given a regular countably additive probability function Pr over a probability space and a random variable u, we define for measurable p, $V(p) = E(u \mid p)$. $V(p)$ is the conditional expectation of u on p. Pr and u are a conceptually coherent prior and utility function pair iff the relation $V(p) < V(q)$ is a coherent initial preference ranking. We shall show that a rational agent with non-trivial preferences must assign two indiscernible propositions the same credence. We begin by showing this for propositions that the agent cares about.

[11] Unfortunately this representation is not unique, except in special circumstances.

Theorem 5.4.4. *Suppose that the agent cares whether p, i.e. $p \prec \neg p$ or $\neg p \prec p$, and the agent's preferences are determined from an expected utility function, V, based on a credence function Cr. Then the agent assigns every proposition indiscernible from p the same credence as p.*

Proof. Suppose that $p \sim q$, i.e., for some $\sigma \in ad(\Gamma)$, $\sigma(p) = q$. By proposition 5.4.2 $p \approx \sigma(p) = q$ and $\neg p \approx \sigma(\neg p)$. Since $\sigma(\neg p) = \neg\sigma(p) = \neg q$ we also have $\neg p \approx \neg q$. Thus we $V(p) = V(q)$ and $V(\neg p) = V(\neg q)$. By inspecting the expected utility equation we can see that $Cr(p)$ can be determined from $V(p)$ and $V(\neg p)$, whenever $V(p) \neq V(\neg p)$ as follows:

$$Cr(p) = \frac{1}{1 - V(p)/V(\neg p)}$$

Thus $Cr(p) = Cr(q)$ □

If p is not $\top$ then for each coherent prior Pr there is some coherent utility function u such that $V(p) > 0$. We therefore get the following corollary:

Corollary 5.4.5. *For any coherent prior Pr, and admissible automorphism σ, $Pr(p) = Pr(\sigma(p))$.*

5.4.1 The Galois theory of Boolean algebras

Many classical theories of vagueness employ a model theory which quantifies over indices which are admissible at [43], or indiscriminable from [138] the actual index. We can put the current model theory in that form too: the index j is admissible at i just in case some admissible automorphism maps j to i. We can in fact reverse the order of definitions. In chapter 1 we argued that the goal of a theory of vagueness should be to elucidate the notion of a proposition being precise (as opposed to characterising the determinacy, or borderliness, operator.) If we instead assume we have been given a set of precise propositions, $\mathbb{B}'$, we can define the notion of an automorphism being admissible.

The assumption we need to make to do this is that negations and arbitrary conjunctions and disjunctions of precise propositions are precise (we argued for this fact in chapter 1.) Given this we may define an admissible automorphism as an automorphism on the algebra of propositions which fixes every precise proposition.

Definition 5.4.5.1. *Let $\mathbb{B}$ and $\mathbb{B}'$ be complete Boolean algebras with $\mathbb{B}'$ a Boolean subalgebra of $\mathbb{B}$. Let G be a group of automorphism on $\mathbb{B}$. We define:*

$$Fix_G(\mathbb{B}) := \{b \in \mathbb{B} \mid \forall \sigma \in G, \sigma(b) = b\}$$

$$Aut(\mathbb{B}'/\mathbb{B}) := \{\sigma \mid \sigma \text{ is an automorphism of } \mathbb{B} \text{ and } \sigma(b) = b, \forall b \in \mathbb{B}'\}.$$

Proposition 5.4.6. *Given these definitions we may show that*

1. *$Fix_G(\mathbb{B})$ is a complete Boolean algebra.*
2. *$Aut(\mathbb{B}'/\mathbb{B})$ is a group.*

Letting $\mathbb{B}'$ be the precise propositions, $\mathbb{B}$ the space of all propositions and G the set of admissible automorphisms we can see how these two notions are interdefineable: we should have $G = Aut(\mathbb{B}'/\mathbb{B})$ and $\mathbb{B}' = Fix_G(\mathbb{B})$.

Does the fact that the set of precise propositions $\mathbb{B}'$ is generated from a group of automorphisms place any constraints on $\mathbb{B}'$ other than ensuring it is a complete Boolean algebra? This prompts a useful definition

Definition 5.4.6.1. *Suppose $\mathbb{B}'$ is a Boolean subalgebra of $\mathbb{B}$. We say that $\mathbb{B}$ is a **Galois extension** of $\mathbb{B}'$ if and only if $Fix_G(\mathbb{B}) = \mathbb{B}'$ where $G = Aut(\mathbb{B}'/\mathbb{B})$.*

One can show that not every extension of a complete Boolean algebra is a Galois extension, although we shall not prove this here. The Galois theory of Boolean algebras is a developed subject so I therefore cannot recite all the relevant results. For further references see the chapter on automorphism groups in the handbook of Boolean algebras vol. 2 [94].

Chapter 6

Propositions and Paradoxes

One of the earliest forms of the liar paradox is known as Epimenides paradox. Here is a version of it

> Everything I have ever said is false. (6.1)

It is often noted that there is a crucial difference between this paradox and the more traditional liar paradox. For unlike the liar paradox, (6.1) does not lead to a contradiction - it is perfectly consistent to treat (6.1) as simply false.

If there is anything paradoxical about Epimenides paradox it is that if someone has said that everything they have ever said is false one can infer *a priori* that they have said something else at some point or other. For if (6.1) is false, then I must have said something true at some point, p. Now since p is true, and the proposition that everything I have ever said is false is not, it follows that I have said at least two things.

In this chapter I shall defend this conclusion and show that it is not paradoxical. In particular, I argue for the claim that:

> Necessarily, if anyone says that everything they've said is false, then they've said something else at some point. (6.2)

What is clearly *not* true, however, is that if I've uttered the sentence 'everything I've ever said is false' I must have uttered some other sentence at some point. The way out, I claim, is to give up the connection between uttering and saying, namely that whenever one utters the sentence 'ϕ' one has automatically said that ϕ.

The other important difference between (6.1) and the liar paradox is that sentences and sentential truth do not play a significant role in (6.1), or the argument for (6.2). Indeed, Prior formulated the paradox in such a way that

(a) It does not involve the truth predicate, as applied to sentences or propositions

(b) It does not mention sentences or appeal to the diagonal lemma,

(c) It does not appeal to any kind of disquotational reasoning,

(d) The conclusion can be proven from what are plausibly logical principles alone.

Since the most prominent solutions to the liar paradox are all directed at a notion of sentential truth and are concerned with disquotational principles for a sentential truth predicate it follows that Prior's paradox is left untouched by standard methods for solving these paradoxes.

On the other hand, however, we shall see that the semantic paradoxes can be construed as a *special case* of Prior's paradox. For just as a person cannot say that they're saying something false, (unless they're not simultaneously saying something else) Prior's argument

shows that no sentence can express the proposition that it expresses something false (unless it expresses two things.)

Prior's paradox places constraints on what one can fear, think, believe and desire. It is widely thought that a theory of 'expressing', which relates the sounds and inscriptions we use to communicate to facts about the world, will ultimately be cashed out in terms of the way the patterns of linguistic usage relate to the propositional attitudes of the people speaking the language in question. In section 6.4 I outline how one might be able to state a theory of expressing in a way that allows the vocabulary used in the account to be applied to itself in all the standard sorts of self-referential ways.

6.1 Prior's paradox

In order to formalise Epiminedes' paradox Prior works within a system of quantified propositional logic. Since Prior seems to be able to derive paradoxical conclusions within the confines of these resources alone – i.e. without a truth predicate, or self-referential sentences, or anything common to the standard paradoxes – it is worth spending some time explaining this system.

Quantified propositional logic adds to the syntax of propositional logic a countable set of propositional variables, $p_i, i \in \omega$, and a special quantifier $\forall$. Each propositional variable is a well formed formulae, and in addition to the standard clauses for building up complex well formed formulae we stipulate that if p_i is a propositional variable and ϕ a well formed formula, $\forall p_i \phi$ is well formed.

It should be noted that quantified propositional logic is an expressively weak fragment of second order logic. In propositionally quantified logic one may quantify into sentence position, whereas in full second order logic one may quantify into predicate position, for predicates of arity ≥ 0, as well as being able to quantify into nominal position (first order quantification.)

Although second order logic has been with us since the birth of modern logic (Frege [46]), it is still regarded as controversial. There are, however, systems closely related to second order logic which are gaining widespread acceptance. One of these is plural logic. Plural logic is modelled on the use of plural quantification in English. Thus, for example, if I ate some Cheerios one can infer that there were some things that I ate. This sentence does not commit us to the existence of anything over and above the Cheerios – in particular it does not commit us to a set of Cheerios which I ate. Neither can such plurally quantified sentences always be paraphrased away in terms of singular quantification over the Cheerios themselves.[1]

I shall employ the expression $\exists xx$ for the existential plural quantifier, and tt for plural terms. While expressively powerful, plural quantification is not genuine second order quantification as Frege envisioned it. In particular plural quantifiers allow one to quantify into the place of a plural term, but not into predicate position. One may achieve something expressively equivalent to monadic second order quantification by introducing a one-many relation $x \prec xx$, which can be read as 'x is one of the xx.' Indeed, in [12] Boolos showed how to interpret monadic second order logic in this kind of plural logic with a $\prec$ relation. One might wonder whether there is anything analogous to genuine monadic second order quantification in English. For example, is there an English equivalent of argument (6.5) which is like (6.4) is to (6.3)?

$$a \prec bb;\ \exists xxa \prec xx \tag{6.3}$$

Felix is one of my cats. Therefore there are some things such that Felix is one of them. (6.4)

[1] See for example, the Kaplan-Geach sentence 'some critics admire only one another.' It can be seen that the plural translation of this sentence can be used to categorically characterise the natural numbers, which implies it cannot be equivalent to any first order sentence.

$$Ga; \exists F Fa \tag{6.5}$$

Felix is a cat. Therefore ??? (6.6)

While plural logic is widely accepted due to the existence of plural quantification in English and other natural languages, it is less often noticed that English does in fact contain genuine second order quantification, i.e. quantification into predicate position, as well. Prior appears to have been the first to notice, prior to Boolos [11], that English contains second order, as well as many other forms of non-nominal quantification.

> We form colloquial quantifiers, both nominal and non-nominal, from the words which introduce questions – the nominal 'whoever' from 'who', and the nominal 'however', 'somehow', 'wherever', and 'somehow' from 'how' and 'where'. No grammarian would seriously regard 'somewhere' as anything but an adverb; 'somewhere', in 'I met him somewhere', functions as the adverbial phrase 'in Paris' does in 'I met him in Paris'. We could also say 'I met him in some place', and argue that people who use such locutions are 'ontologically committed' to the existence of places as well as ordinary objects; but we don't *have* to do it that way. Similarly, no grammarian would count 'somehow' as anything but an adverb, functioning in 'I hurt him somehow' exactly as the adverbial phrase 'by treading on his toe' does in 'I hurt him by treading on his toe.' [Prior, the objects of thought, Ch3 §4]

One can thus complete the English translation of (6.6) as follows: 'Felix is a cat; therefore Felix is something.' Although 'something' usually takes the place of a noun phrase, 'is something' is appearing predicatively in the sentence 'Felix is something'. The argument in (6.6) is not clearly valid if 'something' is being interpreted as a singular quantifier (indeed, the conclusion is grammatical only if 'is' is the 'is' of identity.)[2] Thus it seems that English contains genuine monadic second order quantification as well as plural and singular quantification.[3]

It is less clear, however, whether English contains genuine propositional quantification – i.e. quantification into sentence position. After the paragraph above, Prior goes on to suggest that one could go on to extend English with genuine propositional quantifiers corresponding to the wh-phrase 'whether' just as 'anywhere' corresponds to 'where'. Thus Prior suggests 'if anywhether then thether' for the sentence $\forall p(p \rightarrow p)$.[4]. However, if someone was inclined to deny sense to genuine propositional quantification it is likely they would also deny sense to made up locutions such as Prior's 'anywhether' and 'somewhether'. (For further discussion of the coherence of quantification into sentence position see Grover [53], Küng [76] and Zimmerman [153].)

One might have thought that that quantification into sentence position is possible in English, since, for example, in many cases the complementizer 'that' is optional. Thus for example 'Jane believes that snow is white', 'Jane believes snow is white' and 'Jane believes something' are all grammatical, and on at least one reading the latter sentence involves quantification into a position which can be occupied by a sentence. However, in some cases the complementizer is not optional – for example 'snow is white is a proposition' is not clearly grammatical. Thus if the quantifier 'something' is behaving the same way in 'Jane believes something' and in 'Jane believes something which is a proposition' we cannot truly asimilate this kind of quantification to quantification into sentence position.

[2] I say the alternative reading is not *clearly* valid because some people would say that (6.6) is valid but irrelevant because the conclusion is a logical truth. I would prefer to classify the argument as invalid, on this reading, in virtue of the fact that Felix might not have existed, and that no metaphysical contingencies are logical truths.)

[3] It has also been argued that English can accommodate genuine polyadic second order quantification. This claim see slightly more contentious, but see Rayo and Yablo [114].

[4] See also Prior [107]

So what we are lacking is a clear cut example with quantification in the scope of the complementizer. Prior cites Wittgenstein for a number of putative examples like this, where the expressions 'this is how things are' and 'things are thus' function like propositional variables. See for example

> He explained his position to me, said that this was how things are. (Wittgenstein.) (6.7)
>
> However he says that things are, thus they are. (Prior.) (6.8)
>
> John said that he has said that things are somehow, and thus they are not. (6.9)
>
> John said that however he said that things are, they are not so. (6.10)
>
> No-one can say that however they say that things are, they are not, unless they have said that things are some way that they're not and they have said that things are some way that they are. (6.11)

It should be noted that to express propositions like (6.7)-(6.11) people have often thought it necessary to use a truth predicate. If this way of talking is acceptable, however, no truth predicate is really needed. Examples (6.9) to (6.11) are English formulations of the sentences $s\exists p(sp \wedge \neg p)$, $s\forall p(sp \rightarrow \neg p)$, $s\forall p(sp \rightarrow \neg p) \rightarrow (\exists p(sp \wedge \neg p) \wedge \exists p(sp \wedge p))$ which will play an important role in the discussion to come.

In short, then, the kinds of paradoxes involving propositional quantification I am going to consider are formulable in English, and I think do not depend on aspects of second order logic which do not have any analogue in English. However, since propositional quantification is less widely thought to be as unproblematic as, say, plural quantification or first order quantification, I shall also formulate plural and first order versions of the paradox in section 6.1.3.

It is pretty clear that it is grammatically possible to quantify singularly into a belief context, decomposed in the second way.[5] For this reason I shall also consider a singular version of the paradox in section 6.1.3. This version has the downside that it involves a truth predicate, applied to *propositions*, and appeals to the disquotional schema for propositions and propositional truth.

Throughout this chapter I shall often paraphrase statements involving propositional quantifiers with statements involving singular quantification over propositions. My attitude here is that this harmless provided it is clear that the singular paraphrase really can be said using propositional quantifiers. This is primarily a matter of convenience, employing singular quantification often results in a simpler statement when a strictly correct formulation would be far more complex.

6.1.1 The paradox

Prior's paradox relies on only two premises. The first is simply classical propositional logic, and the second is universal instantiation for the propositional quantifiers.

CL Classical propositional logic.

UI $\forall p\phi \rightarrow \phi[\psi/p]$

and one definition:

($\exists$Def) $\exists p\phi := \neg\forall p\neg\phi$

It should be noted that the principle UI is on its own significantly weaker than standard systems of propositionally quantified logic (see chapter 1.)

[5] I am just making a claim about grammar here, not ontology. One may agree that existential (universal) quantification into a belief context is grammatical even if on thinks it is always false (true.)

Theorem 6.1.1. *(*Prior's Theorem*) The following is a theorem of quantified propositional logic:* $d\forall p(dp \to \neg p) \to (\exists p(dp \wedge p) \wedge \exists p(dp \wedge \neg p))$.

Proof. This is Prior's proof (translated from Polish notation.)

1. $\forall p(dp \to \neg p) \to (d\forall p(dp \to \neg p) \to \neg\forall p(dp \to \neg p))$ (UI)
2. $d\forall p(dp \to \neg p) \to (\forall p(dp \to \neg p) \to \neg\forall p(dp \to \neg p))$ (1, CL)
3. $d\forall p(dp \to \neg p) \to \neg\forall p(dp \to \neg p)$ (2, CL)
4. $d\forall p(dp \to \neg p) \to \exists p(dp \wedge p)$ (3, CL, $\exists$Def)
5. $d\forall p(dp \to \neg p) \to (d\forall p(dp \to \neg p) \wedge \neg\forall p(dp \to \neg p))$ (3, CL)
6. $(d\forall p(dp \to \neg p) \wedge \neg\forall p(dp \to \neg p)) \to \exists p(dp \wedge \neg p)$ (CL, UI, $\exists$Def)
7. $d\forall p(dp \to \neg p) \to \exists p(dp \wedge \neg p))$ (5, 6, CL)
8. $d\forall p(dp \to \neg p) \to (\exists p(dp \wedge p) \wedge \exists p(dp \wedge \neg p))$ (4, 7, CL)

□

Epimenides' paradox, for example, can be generated by interpreting dp as 'Epimenides said that p.' In this case the conclusion reads: Epimenides cannot say that however he says that things are, they are not, unless he has both said that things are some way which they are not, and he has said that things are some way which they are. A more accessible singular paraphrase of this would be: Epimenides cannot say that everything he's said is untrue, unless he's said both a true and an untrue proposition.

The theorem is quite general. d can be substituted for Epimenides thinks (at t) that, fears that, hopes that, believes that, sees that, makes it the case that, brings it about that, wonders whether, queries whether, desires that, supposes that, has evidence that and writes that. Epimenides simply cannot think at t (fear, hope, et cetera) that things are not how he thinks that they are at t, unless he's thinking that things are some other way too.

Another example Prior considers in 'The Objects of Thought' [109] is the paradox of the preface. Here one interprets dp to be 'it is written in the book that p'. Thus, for example, if it is written in the book that something written in the book is false, then it simply can't be the case that everything *else* written in the book is true. Similarly, one cannot believe that at least one of your beliefs is false if every other belief you have is in fact true.[6]

In order to make sense of both these theorems we must make a very crucial distinction between making sounds or shapes which constitute uttering a certain sentence or making an inscription of that sentence, and saying or writing that something is the case. For example, at time t I clearly can *utter the sentence* 'however I say that things are at time t, they are not' whilst uttering nothing else, or inscribe in an otherwise empty book the sentence 'however it is written that things are in this book, things are not so.' And Prior's theorem does not say otherwise. All it entails is that by so uttering you have not thereby said that however you've said that things are at time t, they are not.

Before we move on it is worth noting the scope of the theorem. For example, the theorem says that I can't be thinking at t that things are how I'm thinking they are not at t. Nonetheless, one might argue, this is surely what I *think* I'm thinking in this scenario. And while this might indeed be true, this response has limited applicability, for there are other propositions which I cannot even *think* (mistakenly or not) that I'm thinking. For as Prior points out, if we substitute 'I think that I'm thinking that p' for dp in the

[6]The 'paradox of the preface' usually refers to an epistemological problem, whereas we are concerned with whether you can have the belief in question *at all*. However, at least in the special case where all your other beliefs are true, our question provides an answer to the epistemological problem – for presumably one cannot be required to believe something which it is impossible to believe, no matter how good the evidence that you have a false belief.

theorem, it follows that no-one can think that they're thinking that things are not how they're thinking they think that they are, unless there are other things which they think that they're thinking. Similarly, each instance of the theorem discussed so far might be described as one where I *try* to think, say, bring it about (and so on) that p, but fail. However the theorem applies just as forcefully to 'tries to think (say, et cetera) that p', so there will be propositions one cannot even *try* to think or say, unless one is trying to think or say other things simultaneously.

6.1.2 The significance of Prior's paradox

Prior's paradox is different from the standard semantic paradoxes in a number of ways. For example:

(a) It does not involve the truth predicate, as applied to sentences or propositions

(b) It does not mention sentences or appeal to the diagonal lemma,

(c) It does not appeal to any kind of disquotational reasoning,

(d) The conclusion can be proven from what are plausibly logical principles alone.

One might question (b) on the grounds that some versions involve saying or writing which one could argue relate a person to a proposition by way of a sentence.

In response to this one might note that one can say that snow is white in a number of languages, and even within a language, with a number of sentences, implying there cannot be any particular sentence involved in these paradoxes. Furthermore, even if one could argue that *thinking*, and perhaps even fearing and hoping that snow is white holds by way of some inner language of thought, it is far from clear that seeing that p, or bringing it about that p must hold in virtue of some relation between the agent and a sentence. It thus seems quite a far stretch to blame sentences, and in particular the notion of sentential truth, for these paradoxes.

If this is so, then Prior's paradox stands out from the semantic paradoxes, in that the standard classical response – denying the relevant bit of disquotational reasoning – is not available. Furthermore, a solution to Prior's paradox is not forthcoming from the vast literature on the semantic paradoxes as such approaches deal almost exclusively with sentential truth.[7]

On the other hand, it can be argued that the semantic paradoxes are in fact a special case of Prior's paradox. For this reason I think Prior's paradox is much more than an interesting side issue to the semantic paradoxes, and understanding the propositional paradoxes will help us solve the liar paradox.[8]

The semantic paradoxes are typically formulated with semantic notions such as: truth, satisfaction, denotation and expressing. For the classical logician the culprit principles generating the semantic paradoxes is forced on us by logic alone. One must give up the various forms of disquotational reasoning:

$\ulcorner\phi\urcorner$ is true if and only if ϕ

$\ulcorner\phi\urcorner$ is satisfied by $a_1, \ldots, a_n$ if and only if $\phi(a_1, \ldots, a_n)$

$\ulcorner t\urcorner$ denotes t

$\ulcorner\phi\urcorner$ means that ϕ and means nothing else.

[7] With the exception of Barwise and Etchemendy's [7], most of the standard approaches to the paradoxes have concerned sentential truth, and the standard models for untyped sentential truth predicates just do not clearly transfer to the framework of propositional quantification.

[8] This theme – that the propositional paradoxes are crucial to understanding the liar – is continued in the next chapter where it is argued that we should be classifying propositions not sentences into the paradoxical and non-paradoxical, when it comes to explaining what is wrong with liar-like paradoxes.

However there is no appeal to any such principles in Prior's paradox.

In [106] Prior argues that in fact the first and last paradox can be seen to be instances of his theorem. For any given sentence, s, we may substitute dp for 's is true if and only if p'. Now, through any standard self-referential method, let t be a sentence identical to $\ulcorner \forall p(dp \rightarrow \neg p) \urcorner$, where dp is 't is true if and only if p'. By Prior's theorem we get $d\forall p(dp \rightarrow \neg p) \rightarrow (\exists p(dp \wedge p) \wedge \exists p(dp \wedge \neg p))$ for d under the above interpretation. However the consequent is inconsistent – it says that $Tr(t)$ is materially equivalent to both a true proposition and an untrue proposition. It follows that $\neg d\forall p(dp \rightarrow \neg p)$ – it's not the case that: t, i.e. $\ulcorner \forall p(dp \rightarrow \neg p) \urcorner$, is true iff $\forall p(dp \rightarrow \neg p)$. This is just the negation of an instance of the T-schema.

Similarly we could let t be a sentence identical to $\ulcorner \forall p(dp \rightarrow \neg p) \urcorner$ where dp is read as 't means that p'. Assuming that no sentence can mean both a true and a false proposition[9] it follows that the consequent of Prior's theorem is false, and we get $\neg d\forall p(dp \rightarrow \neg p)$. This is just to say that it's not the case that: t, i.e. $\ulcorner \forall p(dp \rightarrow \neg p) \urcorner$, means that $\forall p(dp \rightarrow \neg p)$.

Under some natural assumptions it can be argued that all of the standard semantic paradoxes can be reduced to Prior's paradox, given that one has the notion of what it means for an open formula, ϕ, to mean that p relative to an assignment. For this fact I refer the discussion to McGee's [91].)[10]

6.1.3 Variations and refinements

There are a number of ways in which to challenge the assumptions of Prior's argument. I shall put aside objections which question the principles of classical propositional logic. The remaining challenges can take two forms. The first kind accepts the coherence of quantification into sentence position but rejects the principle UI appealed to in Prior's proof on the grounds that it implies the existence of impredicatively defined propositions. The other kind rejects the coherence of quantification into sentence position altogether.

I shall address these in turn. I shall show that there are paradoxes which do not rely on the full strength of UI. Then I shall consider other versions of Prior's paradox which do not use quantification into sentence position and instead rely on either plural quantification or singular quantification into attitudes verbs.

The first kind of objection accepts the coherence of propositional quantification but rejects UI on the basis that it implies an impredicative comprehension principle. Impredicativity is thought by Russell [115] to be the source of the semantic paradoxes: the quantifier in the proposition that everything I'm currently saying is false quantifies over itself, which in Russell's view involves a vicious kind of circularity. Russell's proposal is effectively to deny that the proposition that everything I'm currently saying is false exists.[11]

There is on important point I am glossing over: Russell's positive proposal also involves the introduction of new wider ranging quantifier (indeed an infinite sequence of quantifiers) with which one can state, using the new quantifer, that the proposition in question exists. This extra feature of Russell's account will not affect our discussion. Russell's strategy is achieved in this context by restricting UI to its predicative instances

PUI $\forall p\phi \rightarrow \phi[\psi/p]$ where ψ is quantifier free.

Recent philosophers have expressed some sympathy to this kind of solution so it is worth some further discussion (see Kaplan [67], Kripke [75].)

Although it is true that one cannot derive Prior's paradox without impredicative specification, Myhill [96] showed that one could prove a weaker but equally problematic paradox:

[9] Although Prior questions this premise.

[10] Given the expression '$\ulcorner \phi \urcorner$ means that p relative to v', one can define '$\ulcorner \phi \urcorner$ is satisfied by p relative to v' as $\exists p(\ulcorner \phi \urcorner$ means that p relative to $v)$.' Given a satisfaction predicate one can define the notions of truth, denotation as in [91].

[11] Strictly speaking the example I gave is stated in terms of singular quantifiers and Russell's proposal applies to propositional quantifiers.

- $\forall p(dp \leftrightarrow (p \leftrightarrow \forall q(dq \rightarrow \neg q))$

Substituting d for 'I said that', and using a singular paraphrase this shows that I cannot say all on only those propositions materially equivalent to the proposition that everything I said is false.

The proof proceeds by reductio. Assume the claim were true:

1. Assume $\forall q(dq \rightarrow \neg q)$
2. $\forall p(dp \leftrightarrow (p \leftrightarrow \forall q(dq \rightarrow \neg q))$ (assumption)
3. $(p \rightarrow p) \leftrightarrow \forall q(dq \rightarrow \neg q)$ by 1.
4. $d(p \rightarrow p)$ by 2.
5. $d(p \rightarrow p) \rightarrow \neg(p \rightarrow p))$ by 1. and PUI (predicative universal instantiation.)
6. $\neg(p \rightarrow p)$ by 4 and 5. Contradiction.

Therefore we have proved $\neg\forall q(dq \rightarrow \neg q)$, i.e. $\exists q(dq \wedge q)$. Let q_0 be a proposition such that $dq_0 \wedge q_0$ (note that q_0 is available for substitution in PUI.) So

1. dq_0
2. q_0
3. $dq_0 \leftrightarrow (q_0 \leftrightarrow \forall q(dq \rightarrow \neg q))$ by assumption and PUI.
4. $(q_0 \leftrightarrow \forall q(dq \rightarrow \neg q))$
5. $\neg q_0$, contradiction.

The second kind of challenger denies the coherence of quantification into sentence position. Although we cannot therefore apply Prior's original paradox there are other variants which use resources that we uncontroversially have. I shall concentrate on two variant paradoxes: paradoxes using plural quantifiers, and paradoxes using singular quantifiers and a propositional truth predicate.

In order to run the paradox in plural logic we introduce a plural term forming operator $\langle x : \phi \rangle$. However in principle this term forming operator could be eliminated contextually: $\psi \mapsto \exists xx(\forall x(x \prec xx \leftrightarrow \phi) \wedge \psi[xx/\langle x : \phi \rangle])$ (it would make the proofs more complicated.) Note that ϕ needn't contain x free. Here are the relevant principles governing the plural version of the paradox

1. Classical propositional logic.
2. $\forall xx\phi \rightarrow \phi[tt/xx]$
3. $t \prec \langle x : \phi \rangle \leftrightarrow \phi[t/x]$

Let R be any relation

Theorem 6.1.2. *(*Plural version of Prior's theorem*) The following is a theorem of plural logic:* $aR\langle x : \forall xx(aRxx \rightarrow \neg a \prec xx)\rangle \rightarrow (\exists xx(aRxx \wedge a \prec xx) \wedge \exists xx(aRxx \wedge \neg a \prec xx))$.

Proof. Let ϕ be the formula $\forall xx(xRxx \rightarrow \neg x \prec xx)$

1. $\forall xx(aRxx \rightarrow \neg a \prec xx) \rightarrow (aR\langle x : \phi \rangle \rightarrow \neg a \prec \langle x : \phi \rangle)$ (2)
2. $aR\langle x : \phi \rangle \rightarrow (\forall xx(aRxx \rightarrow \neg a \prec xx) \rightarrow \neg a \prec \langle x : \phi \rangle)$ (1)
3. $aR\langle x : \phi \rangle \rightarrow (\forall xx(aRxx \rightarrow \neg a \prec xx) \rightarrow \neg\forall xx(aRxx \rightarrow \neg a \prec xx)$ (3, Defϕ)

4. $aR\langle x:\phi\rangle \rightarrow \neg\forall xx(aRxx \rightarrow \neg a \prec xx)$ (1)
5. $aR\langle x:\phi\rangle \rightarrow \exists xx(aRxx \wedge a \prec xx)$ (1, $\exists$Def)
6. $aR\langle x:\phi\rangle \rightarrow (aR\langle x:\phi\rangle \wedge \neg\forall xx(aRxx \rightarrow \neg a \prec xx))$ (1, step 4)
7. $aR\langle x:\phi\rangle \rightarrow (aR\langle x:\phi\rangle \wedge \neg a \prec \langle x:\phi\rangle)$ (3, Defϕ)
8. $(aR\langle x:\phi\rangle \wedge \neg a \prec \langle x:\phi\rangle) \rightarrow \exists xx(aRxx \wedge \neg a \prec xx)$ (1,2, Def$\exists$)
9. $aR\langle x:\phi\rangle \rightarrow \exists xx(aRxx \wedge \neg a \prec xx)$ (step 7 and 8)
10. $aR\langle x:\phi\rangle \rightarrow (\exists xx(aRxx \wedge a \prec xx) \wedge \exists xx(aRxx \wedge \neg a \prec xx))$ (step 5, step 9)

□

Here R can be any predicate taking a singular and a plural argument. A natural class of interpretations of R is inspired directly from Prior's theorem. If d is any operator we can define $yRxx := dy \prec xx$.

Substituting d for 'thinks that' and a for 'Jane', the upshot of this variant is that if I think that Jane is not among any of the things I think that Jane is among then there are some things which I think Jane is among and she is among them, and there are some things which I think Jane is among and she isn't among them. In particular, I cannot think that Jane is among only those things: there will be some things and some other things, both of which I think Jane is among, but such that there's something that is among the original things but not among the other things.

However other examples do not take the form of Prior's original paradox. For example, paradoxes arise when $yRxx$ is taken to be a non-distributive plural predicate. Suppose that $yRxx$ iff y is thinking about the xx. Then this theorem says that if I'm thinking about those things which are not among any things that I'm thinking about, then I'm thinking about some things which I'm among and I'm thinking of some other things which I'm not among. Therefore I cannot think about the things which are not among any things that I'm thinking about, under that description or any other, unless I am thinking about some other things.[12]

Finally we can run a singular version of the paradox.

1. Classical quantificational logic.
2. $T(that\phi) \leftrightarrow \phi$

Theorem 6.1.3. *(*Singular version of Prior's theorem*) The following is a theorem of first order logic and the propositional T-schema:* $d(that\forall x(Dx \rightarrow \neg Tx)) \rightarrow (\exists x(Dx \wedge Tx) \wedge \exists x(Dx \wedge \neg Tx))$.

Proof. Begin by setting ϕ as $\forall x(Dx \rightarrow \neg Tx)$.

1. $\forall x(Dx \rightarrow \neg Tx) \rightarrow (D(that\phi) \rightarrow \neg T(that\phi))$ (1)
2. $D(that\phi) \rightarrow (\forall x(Dx \rightarrow \neg Tx) \rightarrow \neg T(that\phi))$ (1)
3. $D(that\phi) \rightarrow (\forall x(Dx \rightarrow \neg Tx) \rightarrow \neg\forall x(Dx \rightarrow \neg Tx)$ (2, Defϕ)
4. $D(that\phi) \rightarrow \neg\forall x(Dx \rightarrow \neg Tx)$ (1)
5. $D(that\phi) \rightarrow \exists x(Dx \wedge Tx)$ (1, $\exists$Def)

[12]It is interesting to note that instances of this theorem sounds much less paradoxical when $yRxx$ is substituted for a non-intentional verb. For example: 'this piano cannot be carried by just those things which are not among any things which are carrying this piano'; 'I cannot be surrounded only by those people which are not among any of the people that are surrounding me' sound pretty uncontroversial.

6. $D(that\phi) \rightarrow (D(that\phi) \wedge \neg\forall x(Dx \rightarrow \neg Tx))$ (1, step 4)
7. $D(that\phi) \rightarrow (D(that\phi) \wedge \neg T(that\phi))$ (2, Defϕ)
8. $(D(that\phi) \wedge \neg T(that\phi)) \rightarrow \exists x(Dx \wedge \neg Tx)$ (1, Def$\exists$)
9. $D(that\phi) \rightarrow \exists x(Dx \wedge \neg Tx)$ (step 7 and 8)
10. $D(that\phi) \rightarrow (\exists x(Dx \wedge Tx) \wedge \exists x(Dx \wedge \neg Tx))$ (step 5, step 9)

□

This version of the argument does not invoke quantification into sentence position. It might therefore seem to address the worries of those who deny the coherence of quantification into sentence position. However to run the argument here we must appeal to the proposition T-schema. One might think that this is just as problematic as the sentential T-schema. In fact it is not; it is demonstrably consistent – the consistency is practically a consequence of the possible worlds way of thinking about propositions. We shall discuss this issue further in the next section.

6.2 Operators, predicates, sentences and propositions

For someone who accepts my framework there are only three options they can take. They can either weaken classical logic, restrict propositional universal instantiation (to, perhaps, its predicative instances) or accept the conclusion. I have argued that we should accept the conclusion. However, there are presumably people who do not even accept the framework: who reject the coherence of propositional and plural quantification, and reject the propositional T-schema.

For such people, the only way to make sense of sentences like 'things are however the Pope says that they are' is to paraphrase them in terms of singular quantification over propositions and a propositional truth predicate. One cannot run the argument in this context without helping oneself to the propositional T-schema.

6.2.1 Are sentences and propositions isomorphic?

Propositions, as I understand them, are objective non-representational objects. They are thus very much unlike sentences, which are representational and are only related to the objective world indirectly. The distinction I am after has been put clearly by Ramsey:

> Suppose I am at this moment judging that Caesar was murdered: then it is natural to distinguish in this fact on the one side either my mind, or my present mental state, or words or images in my mind, which we will call the mental factor or factors, and on the other side either Caesar, or Caesar's murder, or Caesar and murder, or the proposition Caesar was murdered, or the fact that Caesar was murdered, which we will call the objective factor or factors; and to suppose that the fact that I am judging that Caesar was murdered consists in the holding of some relation or relations between these mental and objective factors. – F.P. Ramsey, 'Facts and Propositions'.[112]

A natural view about propositions that a hypothetical objector could hold is one in which propositions and sentences are alike in the respects just mentioned. But just because someone thinks that there is some level of representation that lies inbetween public language and the world, doesn't mean they must deny there is any notion of 'proposition' that is purely worldly in Ramsey's sense. There appears, then, to be some kind of verbal dispute about what we call a 'proposition' which I wish to sidestep. Thus, for example,

a representationalist will say that propositions qua the objects which play an important role in semantic theory are not the same as propositions qua denotations of 'that' clauses (see, e.g., Field [38].) I shall henceforth use the term 'proposition' to mean whatever 'that' clauses denote. Singular quantification over propositions in this latter sense is then clearly the nominalised analogue of propositional quantification.

Why might someone reject the T-schema for propositions so understood? A seemingly widespread view is that it must be abandoned for exactly the same reasons that we must abandon the sentential T-schema. It is to this view I now turn.

In fact, there is no straightforward argument that you can generate paradoxes for the propositional T-schema like one can for the sentential T-schema. To see this one can show that the propositional T-schema is completely consistent on a view in which propositions are just sets of worlds, relative to some base model (which can include any number of attitude predicates, whose interpretation can apply to any set of propositions you so chose.) Consider a simple language containing a modal operator, $\Box$, (for distinguishing propositions), a subnective $that\phi$, which forms a term from a (possibly open) formula ϕ, and a predicate T as well as the standard first order logical vocabulary. We can produce a fixed domain S5 model for the T-schema in this language as follows: we let W be any set of worlds, let the domain be $\mathcal{P}(W)$. We then define a non-standard notion of complexity that applies to both terms and formulae in which: atomic sentences and terms not involving $that\phi$ terms have complexity 0, and if ϕ and ψ have complexity n, $that\phi$, $\phi \wedge \psi$, $\forall x\phi$, $\Box\phi$ and $\neg\phi$ all have complexity $n+1$, and if the maximum complexity of terms $t_1, \ldots, t_n$ is n, $Pt_1 \ldots t_n$ has compexity $n+1$. We let the extension of T at a world w be the set of $p \in \mathcal{P}(W)$ such that $w \in p$. We then define truth at a world *and* the denotation of $that\phi$ terms, recursively on the complexity of formulae/terms. This is all standard except that the denotation of $that\phi$ at any world is the set $\{w \mid \phi \text{ is true at } w\}$, which is a good definition since $that\phi$ has greater complexity than ϕ (this is where a version of this proof for sentences would break down.) This validates the T-schema, as well as claims such as $that\phi = that\psi \leftrightarrow \Box(\phi \leftrightarrow \psi)$.[13] This argument does not depend on which attitude predicates we have in the language, or which propositions (sets of worlds) they apply to.

6.2.2 Non-wellfounded propositions

Say that a proposition is *genuinely* paradoxical if, were it to exist, it would be inconsistent with the propositional T-schema. Thus, while the propositions generated by Prior's paradox, and its cousin involving singular quantifiers, are in some sense paradoxical because certain people cannot think or say them, they are not genuinely paradoxical because they are perfectly consistent with the T-schema. The preceding consistency argument shows that one cannot generate genuinely paradoxical propositions from standard resources for constructing propositions.

However one might think that there are sound arguments for the existence of genuinely paradoxical propositions. All one needs, of course, is a proposition which says of itself that it's not true. Surely such a proposition would be easy to generate. For example, consider the proposition that Hector's favourite proposition is not true. Now, since we have not said anything about Hector's taste in propositions, surely it might just turn out that Hector's favourite is *that* proposition. In other words, it might just happen that Hector's favourite proposition the proposition that Hector's favourite proposition is not true. But then we have an inconsistency with the propositional T-schema: the proposition that Hector's favourite proposition is not true is true iff Hector's favourite proposition is not true. However substituting Hector's favourite proposition in the left hand side we get the proposition that Hector's favourite proposition is not true is true iff the proposition

[13] If one wanted to just prove the consistency of the propositional T-schema we could have done with just two propositions, true and false, but we would have still had to go through the non-standard definition of complexity to show that the assignment was well-defined.

that Hector's favourite proposition is not true is not true. This is a contradiction.

Similarly, consider the proposition that the proposition Hector is thinking at t is not true? Now surely, for all we've said, Hector might be thinking *this* proposition at t. Surely he can think whatever he likes, so what could stop him from thinking this proposition? But once we've got this far we can refute the propositional T-schema just as above. Of course, I have denied that Hector can think this proposition at t unless Hector is thinking two or more things at t, but one might prefer to reverse the argument against the propositional T-schema instead.

However good this argument sounds, it runs the risk of over generating. What we have purportedly established is that there a proposition, p, which is identical to the proposition that p is not true. However, the use of truth was not essential to these arguments at all. Consider not the proposition that Hector's favourite proposition is not true, but the negation of Hector's favourite proposition. Call that p. Could it not just happen that Hector's favourite proposition is p, for all the same reasons we mentioned earlier (he can think whatever he wants to think, et cetera)? But if so, then p is identical to its own negation, since p is the negation of Hector's favourite proposition, and Hector's favourite proposition is p.

This is inconsistent with a number of accounts of propositions. If propositions are sets of worlds, or Russellian propositions (assuming a wellfounded notion of constituenthood) then no proposition is identical to its own negation. But even worse, such a consequence seems completely incompatible with classical logic. For example, the proposition that p if and only if $\neg p$ must presumably be true if p and $\neg p$ are identical, but this is simply impossible in classical logic.

If this argument really was convincing, then perhaps we should accept non-wellfounded propositions, and a non-classical logic in which $p \leftrightarrow \neg p$ is consistent. However there are reasons to think that the argument is problematic. Here is an argument for the existence of non-wellfounded numbers. Consider the successor of Hector's favourite number. Call this n. Now since I haven't said anything about Hector's preferences about numbers surely it might just turn out that his favourite number is n. But if this was so, then n must be its own successor, since n is the successor of Hector's favourite number, and Hector's favourite number is n.

As far as I can see all the rhetoric in favour of Hector being able to prefer n over any other number transfers just as easily from the case that his favourite proposition could be p. However, it is clearly not a good argument. Whatever Hector's favourite number is, it can't be the successor of his favourite number, since for all n, the successor of n is distinct from n.

Similarly, whatever Hector's favourite proposition is, it can't be the negation of his favourite proposition. After all, for every proposition p, p is distinct from the negation of p. And whatever Hector's favourite proposition is, it can't be the proposition that his favourite proposition is not true. This again follows from the plausible principle that for all p, p is distinct from the proposition that p is not true.

Much the same can be said for thinking, fearing and other attitudes which can be had towards numbers. For example, I can think of numbers in all kinds of indirect ways. I can think of the number 4, the $50th$ prime number or Hector's favourite number. But I presumably can't think, at a given time t, of the successor of number I'm thinking of at time t, since whatever number I'm thinking of at time t, it's not the same as its successor. Analogously, I cannot think at t the negation of the proposition I'm thinking at t. But there is nothing mysterious about this. Indeed, I can think the proposition I'm thinking at t at some other time or someone else can think this proposition.

6.2.3 Are predicates and operators isomorphic?

Another common thesis in the same ballpark as the thesis that sentences and propositions are isomorphic is the claim that every operator has a corresponding predicate on sentences. Call this the 'correspondence principle.' Since this claim undermines the current strategy of sharply distinguishing between the propositional paradoxes and the semantic paradoxes I shall now consider this view briefly.

It's not completely straightforward to say what it means for an operator to have a corresponding predicate, but let me give a few examples of this claim at work. In order to formalise principles about metaphysical necessity it is common to introduce an operator, $\Box$, and to state principles concerning its logic in this way. For example, provided $\Box$ means 'it's necessary that', we ought to have the following principles, known as 'factivity' and 'necessitation':

$\Box\phi \rightarrow \phi$

If $\vdash \phi$ then $\vdash \Box\phi$.

Now one way to read the correspondence principle is to say that we must therefore have a predicate of sentences, Nec, such that for any sentence, ϕ, $Nec(\ulcorner\phi\urcorner)$ always means the same as $\Box\phi$. Thus in particular, we should be able to always substitute $\Box\phi$ with $Nec(\ulcorner\phi\urcorner)$, and vice versa, so one ought to have also:

$Nec(\ulcorner\phi\urcorner) \rightarrow \phi$

If $\vdash \phi$ then $\vdash Nec(\ulcorner\phi\urcorner)$.

However, it is well known that these two claims are inconsistent (Montague [95].) It follows from the correspondence principle that one must reject the original claims about the logic of the expression 'it's necessary that'. We must either give up factivity or necessitation.

This form of argument can be similarly used to refute some plausible principles about knowledge – namely that knowledge is factive and that someone can know all the classical consequences of factivity. Indeed, factivity isn't even required for these kinds of paradoxes; there are similar paradoxes of belief. A logic of rational belief might involve the following principles

$B(\phi \rightarrow \psi) \rightarrow (B\phi \rightarrow B\psi)$

$B\phi \rightarrow \neg B\neg\phi$

$B\phi \rightarrow BB\phi$

If $\vdash \phi$ then $\vdash B\phi$

It follows by Löb's theorem that the corresponding predicate versions of these principles is inconsistent. Even if it is implausible that these principles apply to any actual person, or even of every rational person, it seems quite hard to swallow that no-one could ever satisfy these principles simultaneously.

I would suggest that these far fetched conclusions are the effects of a bad premise; that every operator (or predicate of propositions) has a corresponding predicate applying to sentences. If the preceding considerations are not reason enough, let me give a proof, within classical logic, that the correspondence principle is false. One instance of the kind of reasoning goes as follows. Assuming the correspondence principle, there can be no operator that obeys the following principles:

$(\phi \rightarrow \blacksquare\phi) \rightarrow \blacksquare\phi$

$(\blacksquare\phi \rightarrow \phi) \rightarrow \phi$

$\phi \rightarrow (\blacksquare\phi \rightarrow \psi)$

Suppose there was. Then by the correspondence principle there would be a corresponding predicate F. Let $\gamma \leftrightarrow F(\ulcorner\gamma\urcorner)$ (*). Then we have $F(\ulcorner\gamma\urcorner)$ by the first principle and the left to right direction of (*), and γ by the second principle and the right to left direction of (*). Thus, by the last principle we have ψ for arbitrary ψ.

So far so good. But here's the rub: if one accepts classical logic, then one already accepts that there is an operator which satisfies these principles, since they are all principles valid when $\blacksquare$ is interpreted as the classical negation operator, $\neg$.

The strategy of blocking Montague's paradox and the knower paradox by insisting on a distinction between operators and predicates of sentences is not only a coherent position, but a position that is forced on us by classical logic. Nonetheless, one might think that there is at least a coherent notion of knowing or believing a sentence (relative to a language) which one ought to be able to make sense of, even if knowledge and belief are not expressed this way in natural language. Perhaps a natural notion would be that a person knows a sentence, s, in the language $\mathcal{L}$, iff however s says that things are in $\mathcal{L}$ she knows that things are that way. This can be expressed in quantified propositional logic as

$$a \text{ knows}_L \ s := \forall p(s \text{ means}_L \text{ that } p \rightarrow a \text{ knows that } p) \tag{6.12}$$

However, even if we accept all instances of the principle 'if someone knows that p then p', we should not expect to be able to derive the principle 'if someone knows-in-English the sentence $\ulcorner\phi\urcorner$ then ϕ' from (6.12), since one cannot assume the principle '$\ulcorner\phi\urcorner$ means-in-English that ϕ' due to Prior's paradox.

6.3 Does the liar express a proposition?

A common reaction to the liar sentence is to respond that it doesn't express a proposition. Of course, this response does not obviously provide a solution to the paradox – for example it does not tell us which principle used in the derivation of the paradox is to blame. However one might think that the thesis that the liar sentence, and similar antinomies, fail to express a proposition offers some kind of diagnosis of the paradox.

A similarly common reaction to paradoxes involves saying that the liar sentence is meaningless. This would follow from the response above if we made the assumption that the meaning of a sentence is a proposition. However they are strictly independent positions; one could, for example, adopt a non-truth-conditional theory of meaning, such as a use theoretic account of meaning, which would allow for sentences to be meaningful without requiring that any sentence expressed a proposition.

However both these views appear to be incoherent (although not formally inconsistent) given the following two principles:

E1. If s expresses a proposition, so does every sentence s implies.

E2. 's does not express a proposition' implies 's is not true'

Suppose that $L =$'L is not true.' By E2. 'L does not express a proposition' entails 'L is not true', i.e. L. If L does not express a proposition then 'L does not express a proposition' does not express a proposition either by E1. and the fact that 'L does not express a proposition' entails L. Similarly if we assume

M1. If s is meaningful, so is every sentence s implies.

M2. 's is meaningless' implies 's is not true'

we can show that if L is meaningless then 'L is meaningless' is meaningless.

But these positions seem to undermine themselves. In the former case the sentence we used to try and express the central claim of the no-proposition view, i.e. the sentence 'L doesn't express a proposition', fails to express a proposition and therefore, presumably, fails to express the central claim. In the latter case the view is, by its very own lights, meaningless. Perhaps it is possible to live with this, as Wittgenstein apparently did [145].[14] However I believe that there are independent reasons to think that the liar sentence expresses a proposition. According to the resulting view the sentence 'L is meaningless' and 'L does not express a proposition' are both perfectly meaningful sentences which express propositions, but are both false. I maintain instead that L, the liar sentence, is both meaningful and expresses a proposition, which I will state using the sentences 'L is meaningful' and 'L expresses a proposition'. The positive view I put forward, unlike its alternative, can therefore be expressed using sentences that are both meaningful and express propositions.

6.3.1 Liars express propositions

Imagine the following scenario, call it scenario 1. Bob and Alice have just pulled a Christmas cracker of the kind that contains a little fact. Bob pulls out the fact card and, realising what it is, utters the sentence 'the sentence written on this card is true' addressing Alice, who he believes doesn't know what the card is. Call this sentence Bob has uttered S. He turns it over and reads

> Cold weather makes fingernails grow faster

Now suppose Alice hears Bob's initial utterance but cannot see what is written on the card. What is Alice's doxastic situation after hearing this utterance? Being a competent speaker of English, and being generally inclined to trust the card, she puts these facts together with the fact that a certain string of letters is written on the card, S, and updates her beliefs. Assuming a standard possible worlds model of belief here is a list of ways in which reading this sentence affects her doxastic state

1. Alice rules out worlds in which snow is white and the sentence on the card is 'snow is green'
2. Alice rules out worlds in which the sentence on the card is 'snow is green'.
3. Alice does not rule out worlds in which the sentence on the card is 'snow is white'.
4. Alice rules out worlds in which the sentence on the card is 'cold weather makes fingernails grow faster' and in which fingernails do not grow faster in cold weather.

How are we to explain facts like 1.-4. above? Firstly note that some facts are more basic than others – 2., for example, can be explained from 1. and the fact that Alice knows that snow is white: since the worlds where snow is not white are already ruled out for Alice, 1. simply amounts to ruling out worlds where the sentence 'snow is green' is the sentence on the card. Fact 1. is but an instance of a regularity among English speakers; not only would Alice rule out these worlds, but any competent speaker of English who believed S had been uttered literally and sincerely. How are more basic facts like 1. explained? A fairly standard assumption is that these facts are explained by appealing to facts about the conventional meaning of S in English. These are facts about sentences and other linguistic items of English that you are required to know in order to be a competent speaker of English. We can distinguish these from general advice about how to communicate well which apply to communication in any language (such as, for example, the Gricean maxims which are not

[14] According to some readings Wittgenstein thought that, although the sentences of [145] are meaningless, they are still illuminating and can be used to point out ineffable truths. It is unclear whether this kind of position fairs better with respect to the paradoxes since the notion of 'pointing out' seems to be subject to the same paradoxes as 'expressing'.

explicitly about English or any any other language.) In other words, we appeal to facts about what the sentence S semantically expresses.

Now imagine another scenario, scenario 2. Scenario 2 is exactly like scenario 1 except when Bob utters S and turns over the card it reads, to his surprise,

> The sentence Bob just uttered is false.

Alice, who as before cannot see the card, is in exactly the same doxastic state as she is in scenario 1. Therefore facts 1.-4. still obtain and require explanation. Just as before, it seems we should explain them by appeal to the conventional meaning of S, i.e., by appealing to the proposition that S semantically expresses.

By a now straightforward argument it follows that S, the sentence on the card or both sentences disobey the E-schema.[15] Without loss of generality, I shall assume that either S or both sentences fail the E-schema; thus we have, in either case, that 'the sentence on the card is true' does not mean that the sentence on the card is true. (Parallel considerations to the ones I'm about to make would apply under the alternative assumption with minor changes to the example.) Now according to the no-proposition view failures of the E-schema like this one arise when the sentence in question does not express anything at all, so S does not express a proposition.

I think these two scenarios raise two related problems for the no-proposition theory. The first problem concerns scenario 2. I shall argue that the no-proposition theorist cannot explain why facts 1-4 hold in this scenario. The second problem concerns scenario 1. Here I show the no-proposition theorist faces a dilemma depending on whether she claims sentences like S express propositions. Either all, or almost all, discourse involving semantic vocabulary is defective, in that it does not express propositions, or the notion of expressing fails to play the right communicative role. No-proposition views sometimes look slightly better when it is *utterances* rather than sentences that express propositions and are the bearers of truth and falsity; this allows different utterances of the same sentence to sometimes express a proposition and at other times fail to express a proposition. Most contemporary no-proposition views are of exactly this type.[16] The problems I raise for communication apply whether or not we take sentence types or sentence tokens to be the bearers of truth and falsity.

Let us begin with the first problem. As we noted Alice will update her beliefs upon hearing Bob's utterance in accordance with 1.-4. Naturally there are plenty of regularities in people's response to hearing utterances of S which can be explained without appealing to the conventional meaning of S – for example, people will invariably come to believe that S has been uttered, or that the utterer has vocal chords, can speak English, and so on. However, what is special about fact 1., for example, is that while the facts just listed would be known to anyone who heard S, whether they spoke English or not, only English speakers would rule out worlds in which snow is not green and the sentence on the card is 'snow is green'. How could we possibly explain this fact without appealing to the idea that in English S semantically expresses a proposition which is incompatible with these worlds?

Perhaps the no-proposition theorist could appeal to the distinction between semantic meaning and speaker meaning introduced in Grice [52]. Another case that is relevantly similar to the one we are considering involves the sentence S' uttered in two different scenarios

> The man drinking martini looks sad

[15] I am retaining the assumption that s is true iff for some p s means that p and p. The example could be adjusted so as not to rely on this assumption by replacing S with the sentence 'the sentence on the card means something true', and replacing the sentence on the card with 'the sentence Bob just uttered means something false'.

[16] For no proposition views that treat utterances as the expressors of propositions see Parsons [99], Simmons [122], Williamson [141] and Gaifman [48].)

In the first scenario I am indicating a man at a party who is drinking a martini, in the second scenario, unbeknownst to me, nobody in the room is drinking a martini, and the man who I am indicating is in fact drinking water from a martini glass. According to the Frege-Strawson view of definite descriptions, the sentence S' expresses a proposition in scenario 1, but not scenario 2 (in order to get a no-proposition view which is perhaps less out of fashion today replace 'the' with 'that' in the above sentence.) Yet it seems that if you heard S' uttered in the situation described, most people would revise their beliefs in a similar manner, whether they were in scenario 1 or scenario 2.

So there is a parallel problem both for the no-proposition view of failed definite descriptions and the no-proposition view about the liar. However it is widely thought that this problem is not devastating to the no-proposition account of definite descriptions (see Kripke [74].) Facts about how one revises one's beliefs when hearing S' can be explained by the distinction between what a speaker means to communicate when they utter a sentence and the sentences literal semantic content. Although in the second scenario described S' does not literally express a proposition, it is clear to a sympathetic audience what the speaker *meant*.

However in order for the audience to work out that Bob didn't mean to say what his utterance literally expressed they must know that the man in question is not drinking martini. In cases where the audience do not know whether the man is drinking martini the audience will come to believe, upon hearing S' from a reliable person, that there is exactly one man in the room drinking martini. I think that this *is* a serious problem for the no-proposition view of definite description (and also for complex demontratives,) and, *mutatis mutandis*, a problem for the no-proposition accounts of the liar. To say that S does not express a proposition in scenario 2 seems to divorce the notion of 'expressing' from its usual role in communication.

Let us turn to the utterance of S in scenario 1. Could we plausibly deny that sentences like S – sentences which are not actually paradoxical but might have been – express propositions? If we ignore the strong intuitive reasons to think S expresses a proposition in scenario 1, the problem with making a general move like this is that it threatens to render almost all of our ordinary talk about truth and falsity meaningless. As Kripke once put it:

> Many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features. Kripke [73]

his point was made using the seemingly innocuous sentence 'most of Nixon's assertions about Watergate are false'. This sentence has truth conditions; it is verifiable by counting Nixon's Watergate related assertions and seeing how many of them are false. The natural thing to say, then, is that whether an utterance of a sentence like S expresses a proposition or not depends on contingent facts which may not be available to either speaker. This also makes it hard to see how we could use sentences like S to communicate effectively.

Our previous discussion makes it clear that on the current view, there can be a regularity among speakers of English to update on a proposition when they hear an utterance of a sentence even when that utterance doesn't express a proposition in English. I take it that if the notion of 'expressing in English' is to have *any* connection with communication in English, then it should at least say that if an utterance does not express a proposition then one *shouldn't* update on any proposition, even if most speakers usually do. However, in general one cannot know whether one is in scenario like scenario 1 or scenario 2, which makes it impossible for a speaker to know when and when not to update on the relevant proposition.

6.3.2 Which proposition does the liar express?

Our thesis, which I take to be established, is thus (E):

(E) L expresses a proposition.

Let us suppose, then, that the proposition the sentence 'L is not true' expresses is p. By Prior's theorem we know that 'L is not true' does not say that L is not true, so we know that whatever p is, it is not the proposition that L is not true. As we will argue in section 6.4.1, a failure of the E-schema is not problematic in and of itself. However it does raise the question: if p isn't the proposition that L isn't true, which proposition is it?

In order to answer this question we should bear two things in mind. The first is that there is no logical reason why L should express any particular proposition; whatever proposition L does express it expresses it only contingently. In order to determine which proposition L expresses, then, we must look to the contingent facts which determine what sentences mean – facts about their usage. The second is that in this regard the proposition that L isn't true is closely connected to the sentence 'L is not true.' Note for example that although L cannot mean that L is not true, the proposition that L isn't true plays the right communicative role with respect to L.

Definition 6.3.0.1. *Say that p satisfies the weak communicative role of s in $\mathcal{L}$ iff all $\mathcal{L}$ speakers know that in any situation in which A and B are $\mathcal{L}$-speakers the following holds*

- *If B knows that A sincerely and knowledgeably uttered s intending to speak in $\mathcal{L}$, B is in a position to know that p.*
- *If A knows that p, then A is in a position to knowledgeably utter s to any $\mathcal{L}$ speaker.*
- *(Optional): These three facts are known to all $\mathcal{L}$ speakers.*

It should be noted that in general multiple propositions will play the weak communicative role for a sentence s. In particular, the proposition that the total number of stars is even, call this p_1, and the proposition that the total number of snowflakes is even, call this p_2, both play this role for the sentence 'the total number of stars is odd'. This is because all $\mathcal{L}$ speakers know that no $\mathcal{L}$ speaker can knowledgeably utter this sentence, and so know that the first condition vacuously holds. Similarly every $\mathcal{L}$ speaker knows that no $\mathcal{L}$ speaker knows p_1 (p_2) and so know the second condition vacuously. For p to satisfy the weak communicative role for s is not in general sufficient for p to be expressed by s, although I suggest that it is necessary. (In the considered example further principles may come into play, such as the claim that of all the properties playing equivalent communicative roles for a single predicate, we should rule out gruesome ones.)

Note, therefore, that the proposition that L is not true plays the right communicative role with respect to L. In particular, note that if B knows A has sincerely and knowledgeably uttered the sentence 'L is not true', then B is in a position to know that L is not true. Why is this? Because every $\mathcal{L}$ speakers knows that A cannot sincerely and knowledgeably utter the sentence 'L is not true' and every $\mathcal{L}$ speaker knows that no-one knows that L is not true. In either case the conditional is trivially satisfied. L does bear a relation to the proposition that L is not true, namely that of having it play its weak communicative role. which would satisfy principles (1)-(4) of the communicative role (although we know that whatever this relation is, it's not expressing.)

So there are two important things to note. Firstly it does not follow that the proposition that L is not true is the *only* proposition that would play the weak communicative role for the sentence 'L is not true'. Secondly, there may be further principles governing expression which go beyond the weak communicative role that would rule the proposition that L isn't true out. For example, consider the principle considered earlier connecting expression with truth:

> If s says that p then s is true if and only if p.

Note that an instance of this principle is: if L says that L isn't true then L is true if and only if L isn't true. From this we may immediately infer that L doesn't say that L is true.

So, to summarise, we now have two facts about the proposition L expresses, p, which will constrain our investigation

1. p plays the same communicative role with respect to L as the proposition that L is not true.

2. L is true if and only if p.

These facts are suggestive. Note the simplest way to get 2. is to substitute p for the proposition that L is true. Since, we may assume, every speaker knows that no-one knows whether L is true, and that no one can knowledgeably and sincerely utter L it follows that on this suggestion p plays the right communicative role. Therefore we might suggest

'L is not true' says that L is true.

Not only does this suggestion satisfy 1. and 2. it offers us a means to explain why it's indeterminate whether L is true or not. According to this suggestion L says that L is true. We might contrast this with the truth-teller sentence – a sentence which also says of itself that it is true. In both cases we can appeal to the plausible principle:

If s says of itself that it is true, then it's indeterminate whether s is true or not.

Since both L and the truth-teller say of themselves that they are true, they are indeterminate.

In the case of the liar, L, we have argued that L expresses the negation of the proposition we were expecting it to. Does this generalise? Could it be the case that for any sentence 'ϕ', either 'ϕ' means that ϕ or 'ϕ' means that it's not the case that ϕ, with the latter case reserved for 'paradoxical' sentences? In order to see why this generalisation fails we need to consider which proposition the following sentence, called S, expresses

(*) The only starred sentence in chapter 7 is not true.

We know, given the fact that S is the only starred sentence in chapter 7, that S does not express the proposition that the only starred sentence in chapter 7 is not true. If we were to apply our previous strategy we might assume that S expresses the proposition that the only starred sentence in chapter 7 *is true* – call this q. Unfortunately q fails both the constraints we placed on L. Firstly, q does not satisfy the necessited version of the principle we appealed to earlier: necessarily, if S says that q then S is true iff q. Why? It seems I could have easily written the sentence 'snow is white' next to the star above instead of S, and I could have done this without changing whether snow is white or changing any of the meaning facts. Intuitively in this situation, S, i.e. the sentence 'the only starred sentence in chapter 7 is not true', would have been false (because 'snow is white', a true sentence, is only starred sentence in chapter 7.) But note that q holds in this situation so we do not have that S is true iff q, even though S says that q. Secondly, q doesn't play the weak communicative role for S. For example, if an English speaker hears 'the only starred sentence in chapter 7 is not true' but has no idea which sentences are and aren't starred in chapter 7, one thing speakers shouldn't rule out based on this information is that the sentence 'snow is black' is starred. The claim that 'snow is black' is starred seems to be perfectly compatible with what S is saying. However q does rule out the possibility that 'snow is black' is the only starred sentence in chapter 7. Therefore hearing S uttered sincerely and truthfully is not sufficient for one to be in a position to know q. By similar reasoning we can argue that knowing q does not put one in a position to legitimately utter S.

What general rules can we extract for working out which proposition a sentence says? We have argued that each meaningful sentence says exactly one thing, and that, as a contingent and empirical fact about English speakers we have the schema: if 'ϕ' says that p then p has the same weak communicative role as the proposition that ϕ (with respect to

the sentence 'ϕ'.) However, as we noted, almost all unknowable propositions have the same weak communicative role with respect to a sentence expressing an unknowable proposition; a stronger constraint which seems to do the work would be:

> If 'ϕ' says that p then p is indiscernible from the proposition that ϕ.

The view is best explained by working through an example. Let $L =$ 'everything L expresses is false'.

1. If 'Everything L expresses is false' expresses only one proposition, it is not the proposition that everything L expresses is false.
2. We have suggested that it expresses the negation of the proposition that everything L expresses is false. I.e. L expresses the proposition that some proposition L expresses is true.
3. Note that if the only thing *d*ed is the proposition that something *d*ed is true, then the proposition that something *d*ed is true is indeterminate. (Prior's version of the truth-teller puzzle.)
4. Since the only thing L expresses is the proposition that L expresses something true, then the proposition that L expresses something true is indeterminate.
5. If p is indeterminate $\neg p$ is indeterminate.
6. The proposition that everything L says is false is indeterminate.
7. The proposition that something L says is true is indiscernible from the proposition that every thing L says is false.
8. 'Everything L says is false' says something indiscernible from the proposition that everything L says is false.

6.4 A theory of truth and semantic expression

How do we communicate using words in a language? What do we need to know about a language in order to successfully receive and convey beliefs using that language? These questions have preoccupied philosophers of language for a large part of the last century. Two versions of Prior's paradox are particularly relevant to this investigation: the paradoxes refuting the following disquotational principles

> 'ϕ' is true in English if and only if ϕ

> 'ϕ' says [expresses the proposition] that ϕ in English

here ϕ may be substituted for any declarative sentence of English. Henceforth I shall use 'says that ... in $\mathcal{L}$' and 'expresses the proposition that ... in $\mathcal{L}$' interchangeably, ommitting 'in $\mathcal{L}$' if no ambiguity is present.

I shall call instances of Prior's paradox like those above 'semantic paradoxes'. In order to properly understand the semantic paradoxes we need to get into difficult questions concerning how we communicate with a language. The main theme of this section will be: disquotational principles are not needed in order to communicate.

By way of doing this it is worth saying a bit more about the notions of 'truth in $\mathcal{L}$' and 'expressing in $\mathcal{L}$' I am intending to capture. I shall begin by saying some true things about expressing, its basic communicative role, which are intended to elucidate the notion to someone who already has it. Let $\mathcal{L}$ be a language, and let A and B range over people who speak and understand $\mathcal{L}$.

1. Every sentence is assocated with a unique proposition, which is the proposition that sentence says in $\mathcal{L}$.

2. If s says that p in $\mathcal{L}$ and B, an $\mathcal{L}$ speaker, knows that A is a trustworthy speaker of $\mathcal{L}$ and knows that A has sincerely and knowledgeably uttered s intending to speak in $\mathcal{L}$, then B is in a position to know that p.

3. If A knows that p, then A is in a position to knowledgeably utter s to any $\mathcal{L}$ speaker.

4. Everyone who speaks and understands $\mathcal{L}$ knows 1., 2. and 3.

We have simplified things somewhat by ignoring context sensitive language; for our discussion the above principles should be enough. One might want also to require something stronger than 4, as Lewis does in [84] for example:

4′. [Lewis-style]: anyone who knows how to speak/understand $\mathcal{L}$ will know 1., 2., 3. and 4′.

The notion of expression I am interested in is supposed to capture the conventional meaning of a sentence. The notion of truth in $\mathcal{L}$ also has an interesting profile of properties relating it to communicative practices, although for now we shall just list the following

5. s is true in $\mathcal{L}$ if and only if for some p, s says that p, and p.

6. If s says that p in $\mathcal{L}$ then s is true in $\mathcal{L}$ if and only if p.

This fixes the notion of truth in $\mathcal{L}$ to a reasonable degree when understood in conjunction with the above principles governing expression. 6. is a consequence of 5. and states a natural way of weakening the T-schema (one can prove the ϕ instance of the T-schema only when one has 'ϕ' says that ϕ.)

These principles are not intended to provide an explicit or implicit definition of 'semantic expression' – that would presumably require a much richer set of principles. However, with these properties in place we can already begin to tell a story about how one communicates in $\mathcal{L}$. If A knows that B understands $\mathcal{L}$, and knows also that B knows it is in their common interest that B knows p, then A knows that uttering s will result in B knowing p. This can be reasoned out from (1)-(5).

Perhaps the most striking feature of all this is that in order to communicate we only need a notion of expressing which satisfies principles (1)-(4). Disquotational principles are not appealed to at any point.

6.4.1 A non-disquotational picture of meaning

Before we ask whether disquotational principles are essential for communication, it is worth saying a few words about the metaphysical and epistemic status of the disquotational principles. The import of the disquotational principles depends on what one means by the words 'true' and 'express.' According to some theorists 'truth' is a purely logical concept whose primary function is to enrich the language in order to allow it to express infinite generalisations. According to this conception of truth the instances of the T-schema that are true are logically true.

According to this use of the word 'true' the T-schema, if it holds at all, is treated as a logical principle. It should be noted that this is not the way I am using the word 'true' in this context. According to my usage sentential truth is a *relation* between a sentence and a language and a context of utterance. In particular, truth is both language relative and context relative: a sentence is true in a language $\mathcal{L}$ and context of utterance c if for some p it says that p, in $\mathcal{L}$ at c, and p. One and the same sentence can say different things in different languages, and can say different things in different contexts.

Furthermore, even when correctly relativised, most instances of disquotational schemata are both contingent and *a posteriori*. For example, take the following instances:

'Snow is white' is true in English if and only if snow is white.

'Snow is white' says that snow is white in English.

Since English speakers could easily have used the word 'snow' to mean grass while keeping the meaning of 'is white' the same, both instances of these disquotational schemata could easily have failed. Furthermore that 'snow is white' means that snow is white in English is an empirical matter – one that a non-native speaker would have to learn in order to speak it. English is not something one can learn *a priori*. The notion of truth being employed here is therefore not a logical one.

This distinction between the logical and empirical conception of truth has methodological consequences when it comes to the question of whether to revise one's logic. Typically, when reasoning about a hypothesis in empirical enquiry, if we discover that the hypothesis leads to a contradiction in classical logic we reject the hypothesis and not the logic. Historically this is the way empirical enquiry has always proceeded. Thus, in particular, if linguistic anthropologists are considering the hypothesis that a linguistic community is using certain symbols to mean a certain sentence is false in their language, and discover that that hypothesis is inconsistent, standard methodology would require them to reject the hypothesis that the community was using those symbols in that way. If the T-schema really is just one empirical hypothesis among many about how the English language is used, the case against classical logic seems to be weaker.

I have thus far argued the disquotational principles are non-logical empirical claims, and that it is easy to conceive of situations in which they are refutable on empirical or logical grounds. But what is the actual epistemic status of the T-schema and the E-schema? The true instances of them are *a posteriori* truths, but are they typically known or easily knowable to competent speakers of English?

Unlike the logical or Tarskian uses of 'true', truth in English is a contingent and use dependent property. Instances of the T-schema for 'true in English' do not express eternal immutable facts.[17] For example:[18]

> It has always been the case that 'Pluto is a planet' is true in English if and only if is Pluto is a planet. (6.13)
>
> It has always been the case that 'Madagascar is an island' is true in English if and only if Madagascar is an island. (6.14)

For example, although Madagascar has always been an island, the *name* 'Madagascar' originally referred to a portion of mainland Africa, so the sentence 'Madagascar is an island' used to be false. We should therefore deny 6.14. Similarly, until recently, the word 'planet' was being used so as to apply to Pluto. However, in 2006 the International Astronomical Union precisified the definition of 'planet' so that it no longer applied to Pluto. While both (6.13) and (6.14) currently hold, they used to be false.

It is easy to imagine that a competent speaker of English could fail to know certain disquotational priniciples. An English speaker alive in the 13th century might have known that the sentence 'Madagascar is not an island' was true in English, since it was at that time, but of course she could not have know that Madagascar was not an island because it was an island. At some point the meaning of 'Madagascar' changed, permitting English speakers to eventually come to know the relevant disquotational principle. Yet it seems unlikely that everyone would come to know this fact; many competent speakers of English would have continued with their old non-disquotational beliefs until they found out the name had changed its meaning.

[17] When I say a fact, p, is eternal I mean only that it has, is and always will be the case that p. Even those who believe that all propositions and are eternally true or false will have some way of making sense of the tense operators I used in the previous sentence.

[18] I owe the second example to Wolfgang Schwarz.

Let us return to our question: are any instances of the T-schema and E-schema stated above required in order for us to be able to communicate? Would we still be able to communicate efficiently had 'snow is white' meant something other than the proposition that snow is white? It seems to me that the answer is fairly clearly 'yes'. Indeed, we can see from past experience that we communicated fine without certain instances of the T and E-schema. Back in the days when 'Madagascar is an island' was false we seemed to get on ok. Note that the failures of the T and E-schema do not imply any kind of expressive limitation on a language. Even back then, it was possible to say, in English, that Madagascar was an island – it's just they wouldn't have used the sentence 'Madagascar is an island' to say this; we would have had to find some other way of referring to Madagascar.

A simple model for meaning

It is quite tempting to think that we can make our sentences mean anything we want them too by changing the way we use them. However, as we have seen with Prior's paradox this is not true in general: no sentence, s, can mean that s means something false (unless s means more than one thing.)

Similarly one might think that we can make any instance of the T-schema true or false just by changing the way we use English. In this section we'll introduce a simple model for thinking about this effect which we'll use later

The basic idea is to utilize a possible worlds model to represent how a sentences meaning and truth can vary from world to world. Among the facts true at a world are meaning facts; facts about how English sentences are used and what they mean at those worlds. This can be represented as a set of worlds, W, and a function, $f : L \times W \to \mathcal{P}(W)$, assigning, for each world and sentence of our language (English, say), a set of worlds: that sentences meaning at that world.[19]

Say that s means p in $\mathcal{L}$ at w iff $f(s, w) = p$, and say that s is true in $\mathcal{L}$ at w iff $w \in f(s, w)$.

Theorem 6.4.1. *Let $\langle W, f\rangle$ be a model and s a sentence. Then at no world w, does $f(s, w) = \{x \mid s \text{ is not true at } x\} = \{x \mid x \notin f(s, x)\}$. In other words, at no world does s mean the set of worlds at which s is not true.*

Note that for cardinality reasons there is always a proposition which s does not mean at any world since there are more propositions than worlds. However the theorem above is a constructive example of a proposition s can't mean. Note that the set $\{x \mid w \notin f(s, w)\}$ is the set Cantor famously used to show that there could be no surjective function form a set to its powerset.

Suppose the language contains a predicate, 'Tr', quote terms 'ϕ', for each sentence of $\mathcal{L}$, and terms L='$\neg Tr(\lambda)$', T = '$\neg Tr(T)$'. The T-schema can be formulated in this framework, and be said to hold at a world w when the following holds:

> $Tr(\text{`}\phi\text{'})$ is true at w if and only if ϕ is true at w.

Theorem 6.4.2. *For any model $\langle W, f\rangle$ and language satisfying the above assumptions we have:*

1. *In no possible world is every instance of the T-schema true. (consider the liar.)*
2. *In no possible world is every instance of the T-schema false. (consider the truth teller.)*

[19] There are a number tricky issues I'm sidestepping for the moment. For example one needs to be careful how one individuates languages – one might think that even if a language doesn't have the interpretation it actually has necessarily, it can't differ *too* much. Furthermore worlds in which $\mathcal{L}$ isn't used at all might be worlds in which sentences of $\mathcal{L}$ don't have any meaning; this possibility is not represented in my model.

3. *There are pairs of instances of the T-schema both of which are possible, but such that their conjunction isn't.*

The last fact represents paradoxes which result from pairs of sentences referring to them. For example, letting S_1 = 'S_2 is not true', S_2='S_1 is true' we can show that the S_1 (S_2 resp.) instance of T-schema is possible: take a world where grass is green and snow is white. Let S_1 mean that grass is green and S_2 mean that snow is black (switch around to get a world in which the S_2 instance of the T-schema holds.) Note that at this world S_1 is true (since it says that snow is white and snow *is* white) and $S2$ is not true (since ...). This entails that S_1 is true if and only if S_2 is not true, which, since S_1='S_2 is not true', just entails that 'S_2 is not true' is true if and only if S_2 is not true. (which is the required instance of the T-schema.)

6.4.2 Can a language contain its own truth predicate?

Many philosophers have asked whether a language can, in some sense, 'contain its own truth predicate' (McGee [91], Herzberger [57].) The purpose of this section is to clarify what this might mean.

Tarski proved a very general theorem that is sometimes taken to bear on this question.

Theorem 6.4.3. *Let $\mathcal{L}$ be any first order language extending the language of Peano Arithmetic and let $\ulcorner\cdot\urcorner$ be a definable Gödel numbering. The schema:*

$$\theta(\ulcorner\phi\urcorner) \leftrightarrow \phi$$

is inconsistent with Peano Arithmetic.

Proof. By the diagonal lemma there is a sentence, λ, such that $\lambda \leftrightarrow \neg\theta(\ulcorner\lambda\urcorner)$ is provable. By this and our assumption we get $\theta(\ulcorner\lambda\urcorner) \leftrightarrow \neg\theta(\ulcorner\lambda\urcorner)$ □

The importance of this theorem is sometimes exaggerated; what we may conclude is that no extension of arithmetic with its own primitive truth predicate can consistently state the T-schema. The point we have been stressing in the previous sections, however, is that the truth predicate (and the expressing relation) are expressions for stating the kind of relation, holding between a language and the objects of belief, that facilitate communication. The primary function of the truth predicate does not involve the disquotational principles, so for all that theorem 6.4.3 says it is an open question whether a language can contain its own truth predicate.

We have a number of arguments that show *for particular languages* $\mathcal{L}$, that $\mathcal{L}$ does not have a predicate expressing truth in $\mathcal{L}$. For example, a simple consequence of Tarski's theorem is that no arithmetical predicate expresses truth in arithmetic.

Definition 6.4.3.1. *An arithmetical formula with one free variable defines a subset of* $\mathbb{N}$, X, *iff* $\phi(\bar{n})$ *is true in the standard model iff* $n \in X$

Theorem 6.4.4. *No formula of arithmetic defines the set of Gödel numbers of true sentences of arithmetic.*

Proof. If ϕ defines $\{n \mid \mathbb{N} \models \psi \text{ and } n = \ulcorner\psi\urcorner\}$ then $\mathbb{N} \models \phi(\ulcorner\psi\urcorner)$ iff $\mathbb{N} \models \psi$, and therefore $\mathbb{N} \models \phi(\ulcorner\psi\urcorner) \leftrightarrow \psi$, which is impossible by Tarski's theorem. □

Assuming, quite plausibly, that an arithmetical sentence is true if and only if it is true in the standard model it follows that no arithmetical predicate expresses truth in arithmetic.

However, there are two things to note about this. Firstly, while truth-in-the-standard-model may in fact be an extensionally correct characterisation of truth in arithmetic it is not intensionally correct: what's true (not true) in the standard model is necessarily true (not true) in the standard model, yet whether a given arithmetical sentence is true is usually

a contingent matter, since we could have used the arithmetical signs differently. In order to have an intensionally correct definition of truth no purely mathematical language will suffice: we will need a language that is able to talk about the way sentences are used in a given population, how the population updates their beliefs upon hearing certain sentences, and so on.[20] Secondly, it is not clear how to generalise the inexpressibility of arithmetical truth in arithmetic to other languages. For example, how might we argue that set theoretic truth is inexpressible in set theory? Tarski's argument that arithmetical truth is not even extensionally characterisable in arithmetic does not generalise to show that there could be no extensionally correct definition of truth in set theory. The reason is simply that there is no such thing as the 'standard' model of set theory; in fact, every model of set theory is non-standard since every model fails to contain at least one set in its domain of quantification (the model itself, for example.)

I want to argue that a language can, in some sense, contain its own truth predicate. However, it is far from clear what the vague expression 'contain its own truth predicate' means. Let us therefore begin by distinguishing various things this could mean:

(i) We may have a language, $\mathcal{L}$, containing a predicate 'P' which expresses the property of being true in $\mathcal{L}$.

(ii) We may be able to have a language which contains a predicate which plays the correct 'conceptual role', interacting in the correct way with other words in the language. This might include its communicative role (as encoded by (1)-(4)), compositional role, and so on.

(iii) We may be able to have a language, $\mathcal{L}$, which contains a predicate which *expresses the same property* as the predicate 'is true in $\mathcal{L}$' expresses in English.

(iv) We may be able to have a language which can express *particular* truths about its own truth predicate. For example, we could have a language, $\mathcal{L}$, containing a connective 'N', and containing a sentence which expresses the proposition that '$N\phi$' is true in $\mathcal{L}$ iff 'ϕ' is not true in $\mathcal{L}$. This is a language which seems to be able to expresse the fact that the connective 'N' expresses negation.

(v) We could have a predicate 'P' that applies to all the sentences which are determinately true in L and doesn't apply to all the sentences which are determinately untrue in L.

Of these claims, (i) is in fact impossible if the language also contains operator expressing negation.

A natural thing to want, if we are aiming for semantic closure, is for our language, $\mathcal{L}$, to have a predicate expressing the property of being true in $\mathcal{L}$. Unfortunately, as I see it, this is just impossible. The reason is quite simple, and does not rely on any kind of disquotational principle.

Let us first consider what it means for a property to be expressible in a language. Let's take a simple example: the property of being a dog. Suppose that the property of being a dog is expressible in a language $\mathcal{L}$, so there is some predicate in the language, 'P', which expresses doghood.

What does this amount to? Well, it surely has the following consequence. Suppose that the language has a name for Fido – the name 't'. Then intuitively the sentence 'Pt' says that Fido is a dog in the language $\mathcal{L}$ (following the logical convention of writing a term next to a predicate to signify predication.) So at the very least the sentence 'Pt' is true in $\mathcal{L}$ if Fido is a dog, and is not true in $\mathcal{L}$ if Fido is not a dog. So we have:

[20]The trick of actualising the truth predicate to make it rigid will not help either. While the equivalence between truth in the standard model and truth will be necessary, it will remain *a posteriori*. It is easy for someone who does not speak the language of arithmetic to deduce that '1+1=2' is true in the standard model, since this is just a mathematical truth, but not at all easy for them to deduce that '1+1=2' is an actually true sentence of arithmetic, since that would require them to know something about the way in which arithmetical symbols are used.

If, in $\mathcal{L}$, 'P' expresses the property of being a dog and 't' refers to Fido, then 'Pt' is true in $\mathcal{L}$ if and only if Fido is a dog.

Unless otherwise stated, 'expresses' shall mean 'expresses in $\mathcal{L}$' and 'refers' shall mean 'refers in $\mathcal{L}$.' This principle is very hard to deny. If either (i) 'Pt' was true when the thing t refers to is not a dog or (ii) 'Pt' is not true when the thing t refers to is a dog, then it is very hard to see how 'P' could be expressing doghood in any reasonable sense. In order for 'P' to express doghood it had better be true of dogs and only of dogs.

This connection between expression in $\mathcal{L}$ and truth in $\mathcal{L}$ is an instance of a general principle. Let $\mathcal{L}$ be a language with first order syntax (this assumption should be dispensible).

PREMISE 1: If the predicate 'P' expresses a given property, F, in the language $\mathcal{L}$, and 't' refers to a in $\mathcal{L}$, then 'Pt' is true in $\mathcal{L}$ if and only if a is F.

Let N be an operator in $\mathcal{L}$.

PREMISE 2: If N expresses negation then '$N\phi$' is true in $\mathcal{L}$ if and only if 'ϕ' is not true in $\mathcal{L}$.

This also seems to be fairly uncontroversial. If this condition is not necessary for an operator to express negation in a language, I cannot imagine what it would mean for a language to contain a negation operator.

Theorem 6.4.5. *No language $\mathcal{L}$ can have the following*

1. *A predicate, P, which expresses truth in $\mathcal{L}$*
2. *An operator, N, which expresses negation.*
3. *A term t such that t refers (in $\mathcal{L}$) to the sentence 'NPt'*

Proof. Suppose for contradiction that there is such a language.

Substituting 'truth in $\mathcal{L}$' for 'F' in premise 1 we get

(*) If 'P' expresses truth in $\mathcal{L}$ then whenever 't' refers to a, 'Pt' is true in $\mathcal{L}$ if and only if a is true in $\mathcal{L}$.

Suppose 'P' expresses truth in $\mathcal{L}$. Since t refers to the sentence 'NPt', it follows by (*) that 'Pt' is true in $\mathcal{L}$ if and only if 'NPt' is true in $\mathcal{L}$.

By premise 2 'NPt' is true in $\mathcal{L}$ if and only if 'Pt' is not true in $\mathcal{L}$. Contradiction. □

On the assumption that English expresses negation and has a liar-like sentence for each of its predicates, we obtain the following corollary

Corollary 6.4.6. *English has no predicate which expresses truth in English. Not even the predicate 'is true in English'.*

To deny premise 1 would be to do violence to our ordinary notion of 'expressing'; so much so as to make the claim that languages can express their own notion of truth no longer interesting. It is possible to have a predicate in $\mathcal{L}$ which applies to some true sentences of $\mathcal{L}$ and doesn't apply to some untrue sentences. But it must always apply to some untrue sentence or fail to apply to some true sentence if the language expresses negation in the sense of premise 2. It is simply too easy to provide predicates that do this. Similar points apply to premise 2.[21]

The last example is worth considering further. There is a certain oddity with corollary 6.4.6. Here we have stated, in English, that English has no predicate expressing truth in

[21] For a precursor of the view I am defending here – that no language can express its own notion of truth – see Herzberger [58]. Herzberger argued that for any language there was a property that language could not express, namely, being heterological in that language (i.e. being a predicate which is not true of itself.)

English: not even the expression 'true in English'. But *in order to state* corollary 6.4.6 we had to use the predicate 'true in English.' Do we even have any guarantee that we've succeeded in expressing what we wanted to express with corollary 6.4.6?

Our aim, therefore, is to show that English (or languages generally) can state particular facts about truth in English, such as corollary 6.4.6, facts about the compositionality and so on, even though, at the subsentential level, it has no predicate expressing truth in English.

6.5 Appendix: A theory of truth and expression

In order to show that the kind of theory outlined in this chapter is consistent we shall spend this section formulating a theory of truth and semantic expression which has the features we've been arguing for, and we shall show it is consistent. In light of our discussion the theory should satisfy a number of constraints:

- The language should be rich enough to talk about its own syntax and prove the existence of self-referential sentences.
- The language should contain an expressing 'connecticate' and a truth predicate with which we can state a number of facts about truth and expressing.
- The language should have propositional quantifiers.
- The language should have a connective for saying when one proposition is indiscernible from another.
- We should be able to state that the sentence 'ϕ' always expresses a proposition indiscernible from the proposition that ϕ.
- We should be able to state in this language various compositionality principles about truth and expressing.
- If ϕ is a theorem of the system then we should be able to express that theorem using the sentence ϕ. The theory should prove that whenever ϕ is a theorem 'ϕ' expresses the proposition that ϕ

For example, we argued for the fifth point in section 6.3.2, but there we offered no guarantee that this proposal would be consistent. And in section 6.4.2 we noted that corollary 6.4.6 appeared to entail that the words we used to state corollary 6.4.6 didn't express corollary 6.4.6. The final point is useful in this respect: we can formulate a version of corollary 6.4.6 within this theory, and the theory also proves that this statement of 6.4.6 in this theory does in fact express corollary 6.4.6.

6.5.1 Axioms

Let $\mathcal{L}$ be the language of arithmetic augmented with

- Propositional quantifiers and a propositional identity connective.
- A primitive connecticate E, taken a term in its first argument and a formula in its second. Exp can be read as saying that x expresses the proposition that p.
- A binary connective $\sim$. $p \sim q$ is read as saying that q is indiscernible from p.

Within this theory we can define a truth predicate $Tr(x) := \exists p(Exp \wedge p)$ and a determinacy operator $\Delta\phi := \forall p(\phi \sim p \rightarrow p)$. We may also define, within the arithmetical fragment of the theory, a bijective Gödel numbering associating each sentence ϕ of $\mathcal{L}$ with a number $\ulcorner\phi\urcorner$. For convenience we shall employ the dot notation: to represent the result of concatenating

'(' with the sentence x numbers with '$\vee$' with the sentence y numbers with ')' we write $x\dot{\vee}y$ (and similarly for other connectives.) We denote the numeral n, i.e. 0 followed by n ′ symbols, as $\mathbf{n}$.

The theory of semantic expression, T, may be stated as follows. I shall write $\vdash \phi$ to mean there is a proof of ϕ in T and $\vdash^* \phi$ to mean there is a proof which does not involve the E axioms.

PA The axioms of PA.

Prop The axioms for propositional quantification and propositional identity.

E1 $E^{\ulcorner}\phi^{\urcorner}\psi \rightarrow \phi \sim \psi$

E2 $\exists p \Delta^n Exp$

E3 $Exp \wedge Exq \rightarrow \Delta^n(p \leftrightarrow q)$

E$\vee$ $Exp \wedge Eyq \rightarrow E(x\dot{\vee}y)p \vee q$

E$\wedge$ $Exp \wedge Eyq \rightarrow E(x\dot{\wedge}y)p \wedge q$

E$\neg$ $Exp \rightarrow E(\dot{\neg}x)\neg p$

E$\forall$ $\forall n Ex[\dot{\mathbf{n}}/\dot{y}]\phi \rightarrow E(\dot{\forall}yx)\forall n\phi$

E4 If $\vdash \phi \leftrightarrow \psi$ then $\vdash Ex\phi \leftrightarrow Ex\psi$

E5 If $\vdash \phi$ then $\vdash E^{\ulcorner}\phi^{\urcorner}\phi$

I1 $\phi \sim \phi$

I3 $\Delta(\phi \rightarrow \psi) \rightarrow (\Delta\phi \rightarrow \Delta\psi)$

I4 $\forall x \Delta\phi \rightarrow \Delta\forall x\phi$

I5 If $\vdash^* \phi \leftrightarrow \psi$ then $\vdash^* \chi \sim \phi \leftrightarrow \chi \sim \psi$ and $\vdash^* \phi \sim \chi \leftrightarrow \psi \sim \chi$

I6 If $\vdash^* \phi \leftrightarrow \psi$ then $\vdash^* \phi = \psi$.

I7 If $\vdash \phi$ then $\vdash \Delta\phi$.

Here Δ^n represents a sequence of n Δ operators. Given these axioms we can derive a number rules and axioms

SC $\exists p \Delta^n(p \leftrightarrow \phi)$

T $\Delta\phi \rightarrow \phi$

Comp$\vee$ $Tr(x\dot{\vee}y) \leftrightarrow Tr(x) \vee Tr(y)$

Comp$\wedge$ $Tr(x\dot{\wedge}y) \leftrightarrow Tr(x) \wedge Tr(y)$

Comp$\forall$ $\forall z Tr(x[\dot{z}/\dot{y}]) \leftrightarrow Tr(\dot{\forall}yx)$

Comp$\neg$ $Tr(\dot{\neg}x) \leftrightarrow \neg Tr(x)$

RT. $(\Delta\phi \vee \Delta\neg\phi) \rightarrow (Tr(^{\ulcorner}\phi^{\urcorner}) \leftrightarrow \phi)$

If $\vdash \phi \leftrightarrow \psi$ then $\vdash \phi \sim \psi$

It is worth remarking upon the restrictions on I5 and I6. The E axioms are not strictly logical axioms, and represent contingent facts about what sentences happen to express. Had the connective '$\neg$' meant something else then E$\neg$ would not have been true. With the unrestricted version of I5 and I6 we would have been able to show that the proposition that $\dot{\neg}x$ expresses the negation of whatever x expresses is identical (or indiscernible) from the proposition that $0 = 0$. But they are not identical or indiscernible; the former is a contingent fact about this language and the former is a necessary truth.

6.5.2 Consistency proof

In order to show that the theory T is consistent we construct a sequence of models $\langle W_n, \mathbb{N}, |\cdot|_n\rangle$ which, firstly, can be shown to model all the axioms and necessitation rules apart from $\mathsf{E5}$ for $n > 1$, and secondly, can be shown to satisfy more and more applications of the rule $\mathsf{E5}$ as n increases.

A model $\langle W_n, \mathbb{N}, |\cdot|_n\rangle$ provides a domain of quantification, namely $\mathbb{N}$, a set of worlds W_n and an interpretation of the expressing relation, $|\cdot|_n$. $|\cdot|_n$ is a function$\mathbb{N} \to \mathcal{P}(W_n)$; roughly speaking $|k|_n$ is the proposition that is expressed by the sentence k is a Gödel number of. Since we ultimately choose W_n to be the set of natural numbers less than or equal to n the connective $\sim$ obtains its interpretation from the numerical structure of W_n.

For each n we construct a structure, $M_n := \langle W_n, \mathbb{N}, |\cdot|_n\rangle$, which will eventually function as a model of $\mathcal{L}$.

$$W_{n+1} := \{i \in \mathbb{N} \mid 0 < i \leq n+1\}$$

$$|\ulcorner\phi\urcorner|_{n+1} = \begin{cases} |\ulcorner\phi\urcorner|_n \cup \{n+1\} & \text{if } M_n, n \Vdash \phi \\ |\ulcorner\phi\urcorner|_n & \text{if } M_n, n \nVdash \phi \\ \emptyset & \text{if } \phi \text{ is not closed} \end{cases}$$

We set $W_0 := \{0\}$ and $|\cdot|_0$ maps everything to $\emptyset$. Given a structure M_n an assignment, v, is a function that takes each propositional variable to a subset of W_n and each singular variable to a member of $\mathbb{N}$. I shall write $v[x]u$ or $v[p]u$ to mean v and u agree everywhere except, possibly, at x/p.

In the definition of $|\cdot|_{n+1}$ we used the expression '$M_n, w \Vdash \phi$'. We must therefore define what it means for $M_n, w \Vdash_v \phi$, read 'ϕ is true at w in M_n relative to assignment v' as follows (suppressing reference to M_n and v where not in use.)

$w \Vdash_v \phi$ if ϕ is an atomic arithmetical sentence that is true in the standard model relative to v. $w \nVdash_v \phi$ otherwise.

$w \Vdash_v p$ iff $w \in v(p)$, where p is a propositional variable.

$w \Vdash E\mathbf{i}\phi$ iff $|i|_n = \{x \in W_n \mid x \Vdash \phi\}$

$w \Vdash \phi \sim \psi$ iff either for all $x \in W_n$, $x \Vdash \psi \Leftrightarrow x \Vdash \phi$ or for all $x, x+1 \in W_n$, $x \Vdash \phi \Leftrightarrow x+1 \Vdash \psi$

$w \Vdash \phi = \psi$ iff all $x \in W_n$, $x \Vdash \phi \Leftrightarrow x \Vdash \psi$.

$w \Vdash \phi \wedge \psi$ iff $w \Vdash \phi$ and $w \Vdash \psi$

$w \Vdash \neg\phi$ iff $w \nVdash \phi$.

$w \Vdash_v \forall p\phi$ iff $w \Vdash_u \phi$ for each $u[p]v$ with $u(p) \in \mathcal{P}(W_n)$.

$w \Vdash_v \forall x\phi$ iff $w \Vdash_u \phi$ for each $u[x]v$

For $n > 1$ M_n provides a model for the theory T minus the rule $\mathsf{E5}$. As n increases M_n satisfies more applications of $\mathsf{E5}$; in particular M_{n+1} satisfies everything provable in T with at most n applications of $\mathsf{E5}$. It follows that every sentence provable from T is true in some model M_n and in every M_k for $k > n$. Since $1 = 0$ is not true in any M_n it follows that T is consistent.

Chapter 7

Revenge

What counts as a 'solution to the liar paradox'? What exactly is it that needs to be solved and what needs to be explained? For the classical logician it is natural to think this involves explaining why disquotational reasoning fails in certain cases when it holds good in so many others, explaining why it is impossible to know whether the liar is true or not, why it is futile to wonder or try to find out which it is, why it is unassertable and perhaps to explain how our intuitions go so drastically wrong.

A standard strategy for answering such questions is to identify a class of sentences as problematic in the hope that restricting attention to sentences outside this class will guarantee that one will not run into trouble. In Tarski's words [131] 'the appearance of an antinomy is [...] a symptom of disease'. Such accounts might identify this disease with being ungrounded (e.g. [73]), possessing a peculiar kind of context sensitivity [99], having been introduced by impredicative [115] or inconsistent [20] definitions, failing to follow analytically from non-linguistic facts [91] or being semantically unstable [54], to name but a few. I shall call all such approaches 'linguistic accounts' in virtue of their identifying the problematic phenomenon present in the semantic paradoxes with some property of the language or linguistic items used to express the paradoxes. In each case sentences said to be in this class of diseased sentences fail to obey the T-schema, are unknowable, are unassertable, and so on, in virtue of having this status.

It is the purpose of this chapter to demonstrate that this strategy is fundamentally misguided. There is no classification of sentences into healthy and diseased according to which ordinary reasoning can be carried out in cases where we are only concerned with healthy sentences. One can show, quite rigorously, that whatever one takes 'healthy' and 'diseased' to mean, so long as your theory states that the T-schema holds for healthy sentences, your theory will prove that its own theorems are diseased. One can turn this around; if your theory of the disease is T, and if T is both consistent and proves that its theorems are healthy then T does not contain the restricted form of the T-schema for healthy sentences. This is of wide significance, I think, as a standard strategy for bracketing the semantic paradoxes in other areas of philosophy is to exclude the sentences regarded as problematic from consideration.

It is interesting that the majority of formal and philosophical approaches to the paradoxes attribute the problematic status to sentences and not to what those sentences say. In the second half of this chapter I lay the foundations for a non-linguistic theory of the problematic phenomenon responsible for the paradoxes. While the linguistic theorist says that certain pathological sentences cannot have disquotational truth conditions, an alernative hypothesis states that certain pathological propositions cannot serve as the truth conditions for the sentence obtained by enquoting. Logically these two hypotheses are very different. Non-linguistic theories of the pathology seem to be very much under explored, yet they appear to be in just as good a position as a linguistic theory to provide a solution to

the paradoxes.[1] One can still describe the conditions under which disquotational reasoning is licensed, and when it isn't, one can outline which assertions are permissible and which aren't, when knowledge is possible, and so on.

7.1 Linguistic theories and revenge

Solutions to the liar paradox usually generate 'revenge paradoxes'; paradoxes structurally similar to the liar involving the vocabulary employed in solving the liar.

Here is a standard recipe for revenge. In the course of providing a solution to the semantic paradoxes the theorist will introduce concepts which help elucidate the kind of phenomenon responsible for the antinomies. Extant examples of this strategy include solutions which appeal to the notion of ungroundedness, semantic instability or impredicativity (to mention only a few.) In doing so the theorist thereby commits herself to the coherence of these notions and to structurally similar self-referential sentences involving them. If done correctly, the revenge paradox will show that the principles putatively governing the concepts employed in the solution commit you to a contradiction. Theories that employ concepts like this are no better than theories that employ the original inconsistent conception of truth governed by the disquotational principles.[2]

This formula for generating revenge paradoxes is somewhat underspecified. Different theorists have different ideas about what a solution to the paradoxes is supposed to do, and therefore appear to have different commitments. It will be worth our while, therefore, to distinguish different kinds of concept which could lead a theorist into trouble. I will concentrate on two.

I take that any solution worth its salt should identify which premise should be abandoned in a given derivation of a semantic antinomy. However it is a common goal among philosophers working on the paradoxes to want more than this – to want a *diagnosis* of what goes wrong in these derivations, and, more importantly, some sort of informative characterisation of the cases where we can and where can't reason naïvely. Charles Chihara calls this the 'diagnostic problem of the paradox.'

> Alfred Tarski once remarked: "The appearance of an antinomy is for me a symptom of disease." But what disease? That is the diagnostic problem. We have an argument that begins with premises that appear to be clearly true, that proceeds according to inference rules that appear to be valid, but that ends in a contradiction. Evidently, something appears to be the case that isn't. The problem of pinpointing that which is deceiving us and, if possible, explaining how and why the deception was produced is what I wish to call 'the diagnostic problem of the paradox'. (p. 590 of 'The Semantic Paradoxes: A Diagnostic Investigation,' Philosophical Review 88: 590-618, 1979) [20]

Russell's theory of impredicative definitions, Kripke's theory of grounding, the Gupta-Belnap theory of circular definitions, McGee's theory of indefinite and definite truth, contextualist theories and Chihara's own inconsistency account of the paradoxes are some, among many, attempts to address the diagnostic problem. In Tarski's metaphor, these theories attempt provide some informative characterisation of the 'diseased' sentences, like the liar, some characteristic which is present in and responsible for the instances of the naïve

[1] One exception to this attitude appears implicitly in Prior's work on the paradoxes see [109], although he doesn't ever develop a theory of what is going wrong in those cases.

[2] One might want to distinguish this strong kind of revenge, which purports to show that the new concepts are inconsistent, from a weaker kind which tries to shows that the solution in question fails to correctly classify the paradoxes the new revenge sentences generate. For example one might introduce a consistent but highly restrictive notion of groundedness which is silent about what makes sentences of the form 's is grounded' grounded or not; while consistent it fails to say anything informative about the sentence which says of itself that it is not grounded and true.

theory which are inconsistent. Whatever form this takes, these theorists are committed to the coherence of a distinction between 'diseased' and 'healthy' sentences. We might call paradoxes formulated using these notions the diagnostic form of revenge.

I shall distinguish this from another form of the revenge paradox which instead focuses on paradoxes closely related to the notion of *assertability*. This is, for example, how Graham Priest presents the revenge paradox[3]:

> There is, in fact, a uniform method for constructing the revenge paradox [...]. All semantic accounts have a bunch of Good Guys (the true, the stably true, the ultimately true, or whatever). These are the ones that we target when we assert. Then there's the Rest. The extended liar is a sentence, produced by some diagonalizing construction, which says of itself just that it's in the Rest. (Graham Priest [103] 'Revenge, Field and ZF'.)

As we shall see later it will be important to distinguish these two forms. It is natural to think that if a sentence is diseased then it's not assertable – in which case it had better not be a consequence of your informal theory. Given this assumption the two forms of revenge amount to pretty much the same thing. While some theorists reject the connection between assertable and diseased sentences (see Feferman [34], Maudlin [89]), we will later show that this distinction in fact *has* to be given up if the disease really is a distinction between sentences. Therefore this second form of revenge presents a problem in its own right.

Let me end by mentioning a paradox that is sometimes discussed under the heading 'revenge': the strengthened liar. The strengthened liar says of itself that it's either false or gappy, where gappy means neither true nor false. Given some pretty uncontroversial logic, even by non-classical standards, this amounts to a sentence which says of itself that it's not true. The 'new' vocabulary needed to generate this paradox is therefore just the negation operator. Since everyone should be able to make sense of negation I do not consider this to present a special problem over and above the problem of giving a consistent account of the semantic paradoxes in a language containing the usual connectives of propositional logic. Sometimes in discussions of three valued logic it is argued that there is some special kind of negation, 'exclusion negation' – having value 0 or $\frac{1}{2}$, which the theorist is committed to but cannot consistently accommodate. Perhaps the middle truth value represents an indeterminate truth status caused by ungroundedness. In this case I think it would be better to assimilate the resulting paradox to the diagnostic form of revenge. Or perhaps one is a paracomplete theorist who explicitly rejects truth value gaps but who nonetheless posit assertability gaps (Soames [125], or Field [40].) Then one might assimilate the objection to the assertability form of revenge. Either way it is not the presence of negation in the language that poses a special problem, but the presence of these other notions.

7.1.1 Revenge paradoxes

In this section I shall discuss various paradoxes that make trouble for certain diagnoses of the semantic paradoxes.

Let $\mathcal{L}^-$ be the language of arithmetic, $\{+, ., 0, \prime\}$ with a function symbol f for each assuming primitive recursive function. We shall take $=$, $\neg$, $\vee$ and $\exists$ as logical constants and will adopt standard definitions of the remaining logical connectives. For each number n, we shall metalinguistically represent the numeral, '0' succeeded by n '$\prime$'s by $\bar{n}$. $\mathcal{L}$ shall represent the language $\mathcal{L}^-$ augmented by the primitive predicates, Tr, Def and H. $\ulcorner\phi\urcorner$ shall represent the Gödel numeral of the formula $\phi \in \mathcal{L}$, relative to some fixed Gödel numbering. The standard 'dot' notation shall be adopted for representing the syntactic operations; for example, I shall write the function taking the Gödel number of a sentence

[3]See also the presentation in [102], §1.7

to Gödel number of its negation $\dot{\neg}$, and so on. If x is the Gödel number for ϕ then I shall write $x[\dot{y}/z]$ for the Gödel number of $\phi[y/z]$. In what follows I shall be considering theories in the language of $\mathcal{L}$ – sets of sentences that are closed under classical logic. I shall consider various principles and rules; I write '$\vdash \phi$', to mean every substitution instance of ϕ should be added to the theory, and 'if $\vdash \phi$ then $\vdash \psi$' to mean that ψ must be in the theory whenever ϕ is in the theory, and similarly for uniform substitution instances of ϕ and ψ together.

The predicates $Def(x)$ and $H(x)$ are to be interpreted according to one's favoured solution to the diagnostic problem. It is assumed for the sake of argument that any satisfactory solution must at a minimum classify sentences into those which are paradoxical, or 'diseased' to use Tarski's metaphor, and those which are not. The predicate H represents the sentences which are alleged not to be paradoxical, for example, 'snow is white' and 'London is the capital of France' – it may thus be substituted for 'grounded', 'semantically stable' or 'definitely true or definitely false'. The predicate Def represents the unproblematically true sentences according to the alleged solution, such as 'snow is white', but not the liar sentence or 'London is the capital of France' – and may be substituted for 'grounded and true', 'stably true' or 'definitely true'. It is very natural to think that Def and H are interdefinable: $H(\ulcorner\phi\urcorner) := Def(\ulcorner\phi\urcorner) \vee Def(\ulcorner\neg\phi\urcorner)$ and $Def(\ulcorner\phi\urcorner) := H(\ulcorner\phi\urcorner) \wedge Tr(\ulcorner\phi\urcorner)$. However, since I can introduce both concepts independently and I at no point use the assumption that they are interdefinable I shall treat them both as primitives. I do, however, use the word 'diseased' to mean, by definition, 'not healthy'. We can simply represent the disease predicate in the language as $\neg H$.

The framework I have outlined allows us to evaluate a number of linguistic approaches to the diagnostic problem by appropriately reinterpreting Def and H. For example.

1. Theories based on grounding, where $H(x)$ may be read as 'x is grounded.' See (Kripke [73], Yablo [149], Leitgeb [79], Maudlin [89].)

2. The revision theory/the theory of circular definitions: $H(x)$ may be read as 'x is semantically stable.' (Herzberger [59], Gupta and Belnap [54], Yaqūb [151])

3. McGee's theory of indefinite and definite truth [91].

4. Inconsistency accounts of the paradoxes (Chihara [20], Yablo [150], Eklund [28], Azzouni [2], Patterson [100], Scharp [117])

5. Contextualist or utterance based theories (Parsons [99], Burge [16], Williamson [141], Gaifman [48], Simmons [122].)[4]

These theories are all linguistic in the sense that it is sentences or utterances, not their truth conditions, which are the bearers of the disease. It is therefore possible to formalise these theories within our current framework. For contextualists who take utterances to be the bearers of healthiness and disease $H(\ulcorner\phi\urcorner)$ should be interpreted as 'every utterance of 'ϕ' is healthy.'

It should be noted that the language $\mathcal{L}$ is surprisingly modest. For most attempts to diagnose the liar paradox the predicate H, on the relevant interpretation, is only the end product. Such theories usually assume a much more substantial vocabulary such as a background language of set theory (see the grounding and revision theories) or philosophical English (such as the inconsistency theories) or a combination of the two. The problem of consistently accommodating all of this further vocabulary would be much harder, and could well leave the theorist open to new paradoxes.[5]

Other interpretations of H and Def are possible, although we shall not consider them here. One would be a non-specific reading, in which H means 'the absence of whatever it

[4]In this last case slightly more care is needed since it is utterances and not sentences that are the bearers of paradoxicality. Nonetheless, we may still distinguish sentences by whether any utterances of that sentence are paradoxical. Therefore $H(\ulcorner\phi\urcorner)$ may be read as 'no utterance of ϕ is paradoxical.'

[5]We see this later with respect to the paradoxes of grounding (see Herzberger [57].)

is that is responsible for the paradoxes.' Another interesting one to consider is 'would not give any one in the set S cause to be concerned about the semantic paradoxes' for various choices of sets of philosophers S. The theorems to follow hold of these interpretations too.

The nice thing about the framework we have outlined is that there are some ready made theorems we can just take off the shelf which place some limitations on what form a solution to the diagnostic problem can take. We shall begin with:

Theorem 7.1.1 (Montague's theorem)**.** *The following are inconsistent in PA*

T. $Def(\ulcorner\phi\urcorner) \rightarrow \phi$

N. If $\vdash \phi$ *then* $\vdash Def(\ulcorner\phi\urcorner)$

The proof of Montague's theorem seems to me to provide a formal basis for the canonical informal formulations of the revenge paradox. One begins by considering a sentence, R, which is stipulated to be identical to the sentence 'R is not definitely true.' Now, the reasoning goes, if R is definitely true then surely we can disquote R and conclude, since R just is 'R is not definitely true', that R is not definitely true. This step is effectively just an appeal to T. Since the assumption that R is definitely true leads to contradiction, we have therefore proven that R is not definitely true. But surely anything we can prove is definitely true, and since we can prove that R is not definitely true, we can conclude that 'R is not definitely true' – i.e. R – is definitely true. This step is just an appeal to N. So we have proven that R is both definitely true and not definitely true. Montagues theorem shows that this informal argument can in fact be made rigorous in the arithmetical setting we have described (for the details see [95].)

While some theorists explicitly deny T (such as McGee [91]) or N (such as Maudlin [89]) it is very hard to shake the feeling of discomfort, and extant explanations of why these principles fail seem far from satisfactory.[6]

Theorem 7.1.2 (Löb's theorem)**.** *The following are inconsistent in PA*

D. $\neg(Def(\ulcorner\phi\urcorner) \wedge Def(\ulcorner\neg\phi\urcorner))$

K. $Def(\ulcorner\phi \rightarrow \psi\urcorner) \rightarrow (Def(\ulcorner\phi\urcorner) \rightarrow Def(\ulcorner\psi\urcorner))$

4. $Def(\ulcorner\phi\urcorner) \rightarrow Def(\ulcorner Def(\ulcorner\phi\urcorner)\urcorner)$

N. If $\vdash \phi$ *then* $\vdash Def(\ulcorner\phi\urcorner)$

Löb's theory significantly weakens T, in Montague's theorem, to D. D merely says that no sentence and its negation are definite. Alternatively, one could replace D with the principle that no contradiction is definite. On readings of 'definite' espoused by inconsistency theorists this principle may not be sacrosanct, but it nonetheless seems very plausible under any other interpretation listed above.

N is as above. Both K and 4 seem to be very natural principles hold if we are reading 'Def' to mean 'grounded truth'. K, for example, is just a formalisation of the way in which a conditional's truth value is calculated according to the strong Kleene scheme on the 'grounded truth' interpretation of definiteness: if a conditional and its antecedent are definitely true, then so is the consequent. D also seems to be motivated in this way since, according to the Kleene scheme, while a sentence and its negation can have value $\frac{1}{2}$, no sentence and its negation can have value 1. 4 seems natural: the fact that 'ϕ' is grounded and true surely directly grounds the claim that '$Def(\ulcorner\phi\urcorner)$' is grounded and true.

Theorem 7.1.3 (McGee's theorem)**.** *The following are* ω*-inconsistent in PA*

D. $\neg(Def(\ulcorner\phi\urcorner) \wedge Def(\ulcorner\neg\phi\urcorner))$

[6] Priest presses these problems against McGee and Maudlin in [105] and [101].

K. $Def(\ulcorner \phi \rightarrow \psi \urcorner) \rightarrow (Def(\ulcorner \phi \urcorner) \rightarrow Def(\ulcorner \psi \urcorner))$

BF. $\forall x Def(\ulcorner \phi[\dot{x}/v] \urcorner) \rightarrow Def(\ulcorner \forall v \phi \urcorner)$

N. *If* $\vdash \phi$ *then* $\vdash Def(\ulcorner \phi \urcorner)$

Here $\ulcorner \phi[\dot{x}/v] \urcorner$, recall, represents the Gödel number of the result of substituting the numeral for the number x denotes for v in ϕ. While the above principles are not strictly inconsistent, they would become inconsistent if you could, as it were, perform infinitely long deductions. The principles stated here are as with Löb's theorem except that we have substituted the 4. axiom for BF. Again, BF. is not plausible on all interpretations of 'Def'. For example, a revision theorist says, ignoring the details for the moment, that a sentence is definite if its truth value eventually settles on a constant truth value along an infinite sequence of models. Now it could be that, quantifying unrestrictedly, for each value x the truth value of ϕ relative to x eventually settles on a constant truth value but there's no point along the sequence at which ϕ's truth value is settled relative to every value x.

Nonetheless, BF seems like the other principles to be motivated if we read Def as 'groundedly true.' BF. is simply a formalisation of Kripke's informal explanation of grounding, in which a universally quantified claim is grounded by each of its substitution instances. Therefore if each substitution instance of ϕ is grounded and true, then so is universally quantified $\forall x \phi$.

These paradoxes seem quite piecemeal with the plausibility of some premisses depending on the particular interpretation of Def. I think the moral to draw is that there clearly is a serious problem in the vicinity for diagnoses of the liar that have this form, even if it is not clear which inconsistency generating principles are crucial. In the next section I shall show that there is in fact one general unified paradox which applies to all the theories consisdered. I shall argue that it places a serious constraint on diagnoses employing a sentential characterisation of the paradoxes.

7.1.2 The impossibility of a revenge free linguistic diagnosis of the liar

I shall take it for granted that the theorist is engaging with the project of providing an informative characterisation of when we can and can't reason naïvely: the project of giving a diagnosis of the semantic paradoxes. The most fundamental constraint on any such theory is that it assert that when a sentence s is disease free we should not be susceptible to the paradoxes. In other words, any adequate theory, T, must assert the sententially restricted T-schema:

SRT. $H(\ulcorner \phi \urcorner) \rightarrow (Tr(\ulcorner \phi \urcorner) \leftrightarrow \phi)$

If 'ϕ' is healthy then 'ϕ' is true if and only if ϕ

SRT. may well be derivable from other principles of one's theory of healthiness and truth, but it might just as well be added as a primitive constraint. Good diagnoses should have SRT: according to the project description, failures of the T-schema are the symptom of the disease we are attempted to diagnose. If the T-schema fails even when we plug in a (supposedly) healthy sentence, we have simply failed to characterise those sentences which cause failures of disquotational reasoning.

Here is the problem: if T is an arithmetical theory which additionally satisfies these two minimal constraints (is closed under classical logic and contains SRT) then T proves that its own theorems are diseased.

Theorem 7.1.4. *Let T be any theory in the language $\mathcal{L}$ containing SRT and Peano arithmetic. Then T proves that its own theorems are diseased (i.e. are not healthy.)*

Proof. In order to show this we construct a sentence, γ, and show that (i) γ is a theorem of T and (ii) $\neg H(\ulcorner\gamma\urcorner)$ is also a theorem of T.

The sentence γ is a familiar revenge sentence, 'I'm either unhealthy or untrue', constructed via diagonalisation.

1. $\gamma \leftrightarrow (H(\ulcorner\gamma\urcorner) \rightarrow \neg Tr(\ulcorner\gamma\urcorner))$ (Diagonal lemma).
2. $H(\ulcorner\gamma\urcorner) \rightarrow (Tr(\ulcorner\gamma\urcorner) \leftrightarrow \gamma)$ by (SRT)
3. $H(\ulcorner\gamma\urcorner) \rightarrow (Tr(\ulcorner\gamma\urcorner) \rightarrow (H(\ulcorner\gamma\urcorner) \rightarrow \neg Tr(\ulcorner\gamma\urcorner)))$ by 1 and weakening biconditional.
4. $H(\ulcorner\gamma\urcorner) \rightarrow \neg Tr(\ulcorner\gamma\urcorner)$ by the classical tautology: $(p \rightarrow (q \rightarrow (p \rightarrow \neg q))) \rightarrow (p \rightarrow \neg q)$
5. $H(\ulcorner\gamma\urcorner) \rightarrow (H(\ulcorner\gamma\urcorner) \rightarrow \neg Tr(\ulcorner\gamma\urcorner))$ by the classical tautology $p \rightarrow (q \rightarrow p)$.
6. $H(\ulcorner\gamma\urcorner) \rightarrow \gamma$ by 5. and transitivity of $\rightarrow$
7. $H(\ulcorner\gamma\urcorner) \rightarrow Tr(\ulcorner\gamma\urcorner)$ by lines 6. and 2. and classical logic
8. $\neg H(\ulcorner\gamma\urcorner)$ by lines 4. and 7.
9. γ by 1., 8. and classical logic.

□

We can turn this into an inconsistency argument against theories which endorse a rule of necessitation for 'healthiness.' Bear in mind, however, that the necessitation rule is much stronger than what we need to prove theorem 7.1.4.

Corollary 7.1.5. *No consistent theory that proves all of its theorems to be healthy can have SRT. among its axioms or theorems. In other words, the theory consisting of SRT. and the following rule of necessitation is inconsistent.*

***H-Nec**: If $\vdash \phi$ then $\vdash H(\ulcorner\phi\urcorner)$*

Before we discuss this paradox it is worth getting clear on what it means for a sentence to be a theorem of one's theory of truth and healthiness. For a sentence to be a theorem of T I just mean that it follows from T's axioms and rules in classical logic. However T itself need not consist only of logical truths. Even unproblematic instances of the T-schema can be contingent; for example 'snow is white' would have been false if we had used 'snow' to mean grass even though snow would have still been white. Theorems of T, such as SRT., therefore, need not be necessarily or logically true. T, as we introduced it, represents the theorist's *commitments*. Theorem 7.1.4 states that any theory which purports to diagnose the liar sentence (i.e. contains SRT) proves itself to be diseased; the upshot is that anyone adopting this theory is committed to accepting sentences which, by their own commitments, are diseased.

In what follows I shall consider the upshot of theorem 7.1.4 when H is substituted for various different readings.

Let us begin with what I'll call the 'ultra-conservative' reading: $H(\ulcorner\phi\urcorner)$ means that the sentence ϕ does not contain the truth predicate. According to ultra-conservativism the sentence "snow is white' is true' is diseased – in fact, even the sentence 'if 'snow is white' is true then 'snow is white' is true' is diseased. Since the latter is a theorem of classical logic it is quite clear why ultra-conservativism states that its own theorems are diseased – we do not even have to consider anything as exotic as γ to see this.

So, according to ultra-conservativism we don't get the necessitation principle H-Nec. However there are two problems with ultra-conservativism. The first is that it does not provide an informative characterisation of the sentences responsible for failures of the T-schema – assuming it in fact excludes all the paradoxes, it excludes more than it needs to,

leaving us in the dark about what the smaller set of sentences that cause the T-schema to fail might look like. The second problem is that it's not even clear that it does the job of excluding all the paradoxes. There are paradoxes involving empirical predicates which ultra-conservatism fails to exclude. For example, suppose that Fred's favourite set is the set of true sentences. If $L =$ 'L is not in Fred's favourite set' then we cannot substitute L into the T-schema, given this empirical fact about Fred's preferences, yet L does not contain the truth predicate.

Another reading of $H(\ulcorner\phi\urcorner)$ that might be worth investigating would be to substitute H for: I would be comfortable substituting 'ϕ' into the T-schema. On this reading SRT amounts, roughly, to the statement that I can recognise which sentences can be disquoted. Say that someone is good at recognising paradoxes if whenever they are comfortable substituting 'ϕ' into the T-schema, then 'ϕ' is true if and only if ϕ. On this reading of theorem 7.1.4 it follows that anyone who is good enough at recognising paradoxes will be too conservative: they will even refrain from substituting into the T-schema logical consequences of the claim that that they're good at recognising paradoxes.

Similarly, some philosophers use the word 'pathological' in a way that would commit them to a version of SRT. It is commonplace for a philosopher to assert a disquotational principle and the follow it with a footnote saying 'of course, I mean to exclude pathological sentences from being substituted into this schema.' For example in [60] Horwich writes 'the axioms of the minimal theory are all propositions of the form 'p' if and only if p – at least, those which do not fall foul of the 'liar' paradoxes'. I think one of the upshots of theorem 7.1.4 is that this strategy for bracketing the paradoxes is not in general safe – Horwich's principle, for example, commits him to sentences which fall foul of the liar paradox.

My real targets, however, are those who want ultimately to diagnose the paradoxes. These theories usually fall into two classes depending on whether they accept the necessitation principle H-Nec. For example in [91] McGee proposes a theory according to which the liar sentence is indefinite. A central feature of indefinite sentences is that they are unknowable and unnassertable. The thought that all your commitments should be definite is captured in McGee's framework by a principle of definiteness introduction – a principle that allows you to infer that 'ϕ' is definite from ϕ.[7] If we substitute $H(\ulcorner\phi\urcorner)$ for '$\ulcorner\phi\urcorner$ is definitely true or definitely false' it is easy to see that H-Nec is valid, so plugging this interpretation into corollary 7.1.5 it follows, given that McGee's theory contains both H-Nec and is consistent, that it does not have SRT. A similar point applies to the theory of semantic stability, endorsed in various forms by Herzberger and Gupta and Belnap [59], [54]. Here we instead substitute $H(\ulcorner\phi\urcorner)$ for '$\ulcorner\phi\urcorner$ is semantically stable'. In [54] Gupta and Belnap show how to consistently add this predicate into the object language. According to their construction the predicate '$\ulcorner\phi\urcorner$ is semantically stable' obeys H-Nec, so as before we may conclude that this theory does not have SRT.

Since theories of this kind believe that their commitments should be healthy, on the relevant readings of 'healthy', they cannot assert SRT. What are we to make of these theories? Consideration of paradigmatically healthy sentences, such as the sentence 'snow is white', reveal that it is clearly *sometimes* ok to apply disquotational reasoning. Once this is conceded it's natural to think that there must be some characterisation of cases that are like 'snow is white' in this respect – this was, after all, the primary goal of a diagnosis of the paradoxes. Yet the theorist under consideration cannot make this further argument, for it would commit her to SRT, a principle which will lead her to make diseased assertions, and ultimately, to make contradictory assertions. Surely a disquotational failure is a sign of disease. Any theory of this disease which cannot assert this basic fact is not a satisfactory theory.

Note that simple variants of theorem 7.1.4 generate problems for theories in which *utterances*, instead of sentences, are the bearers of disease. This is pertinent for contextualist and other utterance based theories such as Parsons [99], Simmons [122], Williamson [141],

[7] See his strong adequacy condition in chapter 8.

Gaifman [48] and others.) Suppose instead that we have a predicate, H, for distinguishing healthy from diseased utterances, a truth predicate Tr for distinguishing true from untrue utterances, and a relation $U(u, s)$ for stating when an utterance u is an utterance of the sentence s. We can then formulate an analogue to the sententially restricted T-schema:

URT. $\forall u((U(u, \ulcorner\phi\urcorner) \wedge H(u)) \rightarrow (Tr(u) \leftrightarrow \phi))$

URT says that every healthy utterance of 'ϕ' is true if and only if ϕ. We can define *sentential* predicates $H'(\ulcorner\phi\urcorner) := \forall u(U(u, \ulcorner\phi\urcorner) \rightarrow H(u)$ and $Tr'(\ulcorner\phi\urcorner) := \forall u(U(u, \ulcorner\phi\urcorner) \rightarrow Tr(u))$. Given URT and the assumption that there is at least one utterance of ϕ we can derive a version of SRT. involving H' and Tr'. The upshot of theorem 7.1.4 on this interpretation is that (i) the relevant instance of γ is a theorem (γ says, roughly, that not every utterance of it is healthy and not every utterance of it is true) and (ii) if there's an utterance of γ there's an unhealthy utterance of γ.

For most utterance based theories of truth the relevant notion of unhealthiness is that of the utterance expressing a proposition. I shall focus on this interpretation of H for concreteness' sake, although nothing turns on this. Utterance based theories will often make assertive utterances of sentences even when there are other utterances of that sentence which do not express propositions. For example if u is an utterance of the sentence 'u is not true' then the theorist will make utterances of this sentence even though u itself does not express a proposition.

The problem I am describing, however, is much worse than this: it shows that even when the utterance theorist is *describing* her view she can end up failing to say anything. The theories under consideration endorse the principle that when an utterance of 'ϕ' *does* express a proposition it expresses the proposition that ϕ; so they therefore endorse URT. Now imagine a theorist who for the first time is considering the sentence γ. We note that γ is just a logical consequence of URT and the claim that there is at least one utterance of γ. She ends her proof of theorem 7.1.4 by uttering γ for the first time ever, and as it happens, the last time (we can suppose the world ends after she finishes the proof.) Since this is the only utterance of γ, and we've shown that some utterance of γ doesn't express a proposition, it follows that this utterance does not express a proposition. Either the theorists initial utterance of URT, and the following utterances she performs while making the deduction, fail to express propositions, or at some point during her proof of γ her utterances stopped expressing propositions.

Paradoxes for utterance based theories are often worse than the one described above, since they typically endorse something stronger than URT. Consider the following argument for an instance of the T principle in Montagues theorem, where $Def(\ulcorner\phi\urcorner) = \exists u(U(u, \ulcorner\phi\urcorner) \wedge Tr(u))$:

> If an utterance is true then it says something.
>
> If an utterance is true and says that p then p.
>
> If an utterance of 'ϕ' says anything at all it says that ϕ
>
> Therefore: if some utterance of 'ϕ' is true, then ϕ

If L = 'no utterance of L is true', then by the above fact it follows that if some utterance of L is true, no utterance of L is true. Thus, by reductio, it follows that no utterance of L is true. Anyone who endorses these two premises is committed to the claim that no utterance of L is true, and therefore should surely be able to state the fact that no utterance of L is true. Yet when the theorist tries to express this commitment, by uttering the sentence 'no utterance of L is true' she fails. Since according to her own view there are no true utterances of the sentence 'no utterance of L is true' – even her very own utterances fail to be true; presumably by failing to express a proposition.

Let me end my discussion by considering those who explicitly disavow the connection between assertable and diseased sentences (see for example Feferman [34], Maudlin [89].) The interest of theorem 7.1.4 is rather that adopting a linguistic diagnosis of the paradoxes *forces* us to prize these two notions apart. For those who already distinguish them, it is hardly a troublesome consequence.

As it happens, those who identify the disease with ungroundedness tend not to worry about asserting ungrounded sentences. For example, they will typically assert 'the liar is neither true nor false', a sentence which is, by their own account, both untrue and ungrounded ([89].) While they do not appear to fall afoul of this theorem, let me say a few things about this.

Firstly there is nothing about the notion of ungroundedness itself which makes plausible the claim that it is OK to assert ungrounded sentences. In general, then, there is a challenge for this kind of theorist to explain what they're doing when they assert. Only Maudlin has attempted to do this ([89] chapter 5), but as Priest objects [101], the theory of assertability seems quite arbitrary; you're not allowed to assert an ungrounded sentence if it is atomic, but you may assert an ungrounded sentence if it is the negation of an atomic sentence. In Priest's words 'truth is the aim of assertion. Once this connection is broken, the notion of assertion comes free from its mooring, and it is not clear why we should assert anything.'

Secondly, and more importantly, there are other revenge paradoxes which these theorists do not escape. In distinguishing between assertable and diseased sentences they distinguish the diagnostic form of revenge from the assertability form of revenge outlined at the beginning of this section. In making this distinction they are forced to treat both paradoxes separately. Maudlin attempts to address these in [90], but does so at the cost of giving up his theory of assertability.

As well as the assertability form of revenge, there are paradoxes of grounding which do not rely on the principle of necessitation. Herzberger shows that there are sentences, s, which are grounded by all and only the grounded sentences; for example he suggests that the sentence 'every grounded sentence is true or false' is grounded by all and only the grounded sentences. But (i) if everything that grounds s is grounded then s is grounded and therefore, (ii) since s is grounded by *all* grounded sentences s grounds itself. It follows that s grounds itself and is thus ungrounded after all, a contradiction. This known as Herzberger's paradox [57], which is closely related to Mirimanoff's paradox [93].[8]

7.2 Non-linguistic theories

In the last section I outlined some very general problems. Now I want to sketch what I think is the most promising avenue for avoiding these paradoxes. The key idea involves rejecting the assumption that the distinction between the good and the bad instances of the T-schema must be grounded in a distinction between the kinds of sentences we can substitute into the left hand side of the T-schema. The T-schema associates sentences with truth conditions; on the opposing view it is rather the truth conditions, not the sentences, that must be vetted for disquotational reasoning. In order to properly understand this it is crucial to have two assumptions out in the open. These are:

CL. Classical logic.

CS. If s says that p then s is true if and only if p.

[8]See also, for example, Leitgeb [79] Lemma 14.9: 'There is no sentence $\phi \in \mathcal{L}_{Tr}$ s.t. ϕ depends on Φ_{lf} essentially (i.e. on the set of sentences that depend on non-semantic states of affairs).' Leitgeb uses this to show that one cannot define groundedness, i.e. Φ_{lf}, arithmetically. However it seems to show that one couldn't even define the set of grounded sentences set theoretically at all. Leitgeb, in personal communication, has further told me that he does not think that one could accommodate a binary grounding relation in one's object theory, even though it may be possible to have a groundedness predicate in the object language. See also McGee, theorem 5.11, which also seems to suggest that one cannot formulate an informative definition of grounding in the object language.

Classical logic should be self explanatory. CS, on the other hand, encodes a principle (practically a definition) of classical semantics: that what a sentence says (or means or expresses) is its truth conditions. If one can make sense of quantification into sentence position then the expression 'true' can be defined in terms of 'says that': s is true just in case for all p if s says that p then p. If one made a standard assumption in classical semantics, namely that every sentence means at most one thing, then one can prove CS from this definition. Indeed, one needs only the substantially weaker schema: if s says that p and s says that q then p if and only if q. CS is partly substantive – it contains the assumption that every sentence says at most one thing – and partly verbal – it is a matter of choosing to use the words 'true' and 'says that' in a related way.

While it is well known that classical logic proves that we cannot have all instances of the T-schema,

T. 'ϕ' is true if and only if ϕ

it is less often observed that the disquotational saying schema

S. 'ϕ' says that ϕ

is also problematic. To my knowledge this was first shown by Prior in [109] (see chapter 6.) Prior's argument required one to be able to quantify into sentence position and required the assumption, mentioned above, that every sentence says at most one thing.[9]

In fact you can refute S directly from CS without invoking anything as controversial as quantification into sentence position. First note that we can prove that no sentence can say of itself that it is not true. For if s said that s is not true then by CS s would be true if and only if it is not true. Secondly note that if we considered the sentence L='L is not true', the saying schema, S, would entail that L says that L is not true.

Much of what I say about sentences and truth conditions will make little sense if the reader assumes they are related by schemata such as S or T. It is absolutely crucial in what follows to bear in mind that neither T nor S hold unrestrictedly.

7.2.1 Unclear truth conditions

Despite the paradoxes we have just discussed, it is hard to deny that there is a distinction between two kinds of sentence. Some sentences are clearly true, 'snow is white' for example, and others are neither clearly true nor clearly false, such as the liar sentence or the truth teller. None of the paradoxes we consider cast doubt on the existence of such a distinction. However, whatever clear truth is we know it cannot satisfy both SRT and a necessitation principle.

Taking this proposal at face value, 'clearly' does not function as a predicate, and it is not sentences that are clear or unclear. In 'S is clearly true', 'clearly' functions as an adverb, modifying the truth predicate just as 'possibly' does in 'S is possibly true', and 'not' in 'S is not true'. In philosophical discourse we often regiment adverbial modification by way of a sentential operator, so that for example 'Hector is a possible skater' can be regimented as 'it's possible that Hector is a skater'. This way we may apply clarity, possibility, negation, and so on even to sentences that are not in subject predicate form.

'True' is not the only predicate that 'clearly' modifies, thus while clear truth certainly is linguistic, in some sense, clear tallness, say, need not be. Consider

1. Snow is clearly white.

2. Clearly, snow is white

[9]The premises of the argument do not preclude the possibility that some sentences don't say anything. Although Prior doesn't mention this explicitly, it actually only requires the assumption that each sentence only says materially equivalent things.

Neither 1, nor its regimentation using an operator expression, 2., seems to be about language in any way. Regarding tense operators Prior wrote:

> When a sentence is formed out of another sentence or other sentences by means of an adverb or conjunction, it is not *about* those other sentences, but about whatever they are themselves about. – [108], "changes in events and changes in things"

Unless p itself happens to be about language, when we say that it's clear that p we are not talking about language any more than we would be if we said it's necessary that p. In the contexts we will apply this distinction, however, p will often concern language so it pays to be especially careful about the distinction. Consider

3. 'Snow is white' is clearly true

4. Clearly, 'snow is white' is true.

While it seems that 1 is equivalent to 3 and 2 is equivalent to 4, 3 and 4, unlike 1 and 2, concern English sentences and they are at best contingently equivalent, since we could have used 'snow is white' to mean that Harry is bald, for example. Even if 'snow is white' had meant something unclear, snow would still be clearly white.

I hope to have introduced this notion in such a way that we may have a neutral understanding the notion without having a theory of it. Anyone who has thought about the paradoxes enough to recognise the difference between the claim that snow is white and the claim that L is true, where $L =$ 'L is not true', should be able to understand the distinction. And the regimentation of this distinction using an operator expression seems reasonable given that other non-theoretical ways of introducing the distinction involve some kind of adverbial modification of 'true'. For example, we might said that 'snow is white' is definitely true, straightforwardly true, determinately true and so on.

Unclarity, on this picture, is not something sentences possess but something their truth conditions possess. If I say that Harry is bald then it is what I have *said* that is unclear, not the sentence I used to say it. With this distinction in hand a subtle assumption being made in diagnoses of the naïve theory of truth is revealed. Consider

'ϕ' says that ϕ in English

'ϕ' is true in English if and only if ϕ

The left hand part of both of these equations concerns language, whereas the right hand parts concern the world in some sense; together they match sentences with their truth conditions. The assumption being made is that we shouldn't accept instances of these schemata which involve sentences with certain disquotation preventing features; for example we should not accept instances in which the left side involves an ungrounded sentence, or a semantically unstable sentence. That is, we are placing a 'language side' restriction as opposed to a truth conditional restriction.

But what of the other possibility? Might there be truth conditions which are not suitable for standing in an 'enquotational relation' to a sentence? Could it be that unclear propositions are incapable of being easily said by certain sentences of a given language, and are unsuitable for being their truth conditions?

Let me begin by listing some possible reasons why a proposition might be unsuitable to be expressed by either a certain sentence or any sentence in a language. (1) If, for example, it is common knowledge among speakers of English that no English speaker knows whether p, could there be any conventional link between the dispositions to utter a sentence s and our knowledge of p?[10] Yet one might think that unclarity involves just this kind of feature.

[10]See Lewis [83] footnote 4.

(2) If p entails things about s's truth or untruth (i.e. things about whether s means a true or false proposition) p may be unsuitable as a truth condition for s because it may be literally metaphysically impossible for a linguistic community to use s to mean that p, since it is impossible to use s to a mean a proposition with a different truth value to the proposition s means. (3) On a correspondence theory of sentential truth if p lacks a truth maker then there is no fact for s to correspond to; this could be considered sufficient to prevent s from saying that p. (4) Some propositions concerning meaning facts are false at nearby worlds due to semantic plasticity. Propositions whose truth depends on variations in the way we use language may be related to English in a way that makes this relation unsuitable to be stated using a disquotational principle since the relevant disquotational principles may be *unsafe*.

I do not want to adjudicate between these possible explanations of the unsuitability of certain instances of the disquotational schemata; I merely want to point out that there are no shortage of explanations which identify the truth conditions, rather than the truth bearers, as culprits for the inconsistency of the naïve theory of truth. So it is of great relevance that substituting any of the above constraints into the standard revenge arguments do not motivate any theory with the form of SRT which tells us when a sentence can have its disquotational truth conditions:

SRT. If 'p' is clearly true or clearly untrue then 'p' is true if and only if p.

They motivate instead principles which tell us when the truth conditions for a sentence are 'enquotational':[11]

RT. When it is clear that p or it's clear that not p, then 'p' is true iff p.

RS. If it's clear whether or not p then 'p' says that p.

When theorising about the liar one cannot be too careful about use and mention. Here the pedanticism pays off: RT and RS are not equivalent to SRT or a sententially restricted version of RS, they are enquotational in nature rather than disquotational. The most one can prove from RT., for example, is that if a sentence 'p' is clearly true or clearly untrue then "'p' is true' is true if and only if 'p' is true. Compare this, for example, with the instance of SRT considered in the last paragraph.

What other principles should we expect to govern the notion of clarity? One constraint, which we motivated in our discussion of assertability in the last section, is that it is in general bad to believe and assert unclear propositions. The necessitation rule, Nec., ensures that everything the theory commits you to is clear.[12] What is clear is closed under classical logic, so we should endorse the closure principle for clarity, K, below. Finally note that the informal notion of clarity is factive, so we add the schema T. The resulting modal system is called KT

K If it's clear that if p then q and it's clear that p then it's clear that q.

T If it's clear that p then p.

Nec. If you can prove that p, then you can prove that it's clear that p.

[11] It should be noted that RS runs into trouble with contingent liars. For example, it may just happen to be clear that Alice's favourite sentence is true because her favourite sentence is 'snow is white'. Thus RS commits the sentence 'Alice's favourite sentence is true' to expressing a proposition it cannot express had Alice's preferences been different. But it's very natural to think that if we are to communicate successfully what a sentence says should not depend on facts ordinary speakers have no access to, such as details concerning Alice's preferences between sentences.

[12] Although it is strictly stronger: it ensures that everything the theory commits us to is provably clear. This may be too strong since it commits us not only to all the axioms being clearly true, but also to them being clearly clearly true, and so on for any number of iterations of 'clearly'.

Call the logical system you get by adding RT to KT: BCT – the basic logic of clarity and truth. BCT looks like it should be susceptible to versions of Montague's paradox and theorem 7.1.4. However it is not: BCT is a consistent theory (BCT is a consistent fragment of the theory whose consistency we prove in theorem 7.3.1.)

Given this fact it is of course natural to wonder what happens with the revenge sentences. Let's introduce the name R for the sentence 'R is not clearly true'. The first thing to observe is that the standard way of introducing the revenge problem, described below, rests on a confusion of use and mention in the present framework:

> Suppose that R is clearly true. Since we can disquote clear truths, and since R='R is not clearly true', it follows that R is not clearly true. This contradicts our assumption.
>
> So R is not clearly true. But we've just proved the sentence 'R is not clearly true', which is just R, so R is clearly true. Contradiction.

The mistake in this argument is found in the line 'we can disquote clear truths'. We may write 'R is clearly true' more perspicuously as 'clearly, R is true'.[13] What is clear, as already stressed, is not a sentence, R, but the proposition that R is true, and clarity only licenses the proposition that R is true to serve as the truth conditions for the sentence 'R is true' – not for the proposition that R is not clearly true to serve as the truth conditions for the sentence 'R is not clearly true'. The mistake was to disquote R (= 'R is not clearly true') when the correct principle is an enquotational one: that R is true if and only if 'R is true' is true.

The most natural way to modify the argument so that we can apply RT is to instead assume that it's clear that R is not clearly true (i.e. instead of assuming that 'R is not clearly true' is clearly true.) The argument would then proceed as follows

> Suppose that it is clear that R is not clearly true. Then, of course, R is not clearly true. But since we can enquote in clear cases, we also have 'R is not clearly true' is true. 'R is not clearly true' just is R so R is true. So we have that R is true but not clearly true.

Is this a contradiction? No, we know that there are unclear truths: either the liar is true or it isn't but either way it is unclear. Of course, no-one is forced to endorse the claim that R is true but not clearly true on the basis of this argument – the point is that nothing contradictory follows from the assumption that it is clear that R is not clearly true. It seems, therefore, that once we are careful about use and mention and are careful to distinguish enquotational principles from disquotational principles, then the revenge paradox does not get off the ground.

7.2.2 Higher order unclarity

In the previous section we introduced a distinction between clear and unclear cases of truth. One of the benefits this way of drawing the distinction is that 'clearly' is not a predicate in its own right and can just as well be used to state that someone is clearly bald as it can be to state that a sentence is clearly true. This freedom allows us to formulate questions about iterations of the clarity operator. We argued that the distinction between being true and being untrue is not always a clear one, as the liar paradox demonstrates. A natural follow up question concerns the distinction between being a clear truth and an unclear truth: is

[13] The rest of the argument is in fact good by the lights of BCT. In particular the move from proving that R is not clearly true to proving that 'R is not clearly true' is clearly true is fine, although not obvious. If we have a proof of p, then by Nec. we can prove that it's clear that p. By RT. 'p' is true iff p, and since we have p we have 'p' is true. Finally by Nec. again we can prove that 'p' is clearly true.

this distinction a clear one? The liar paradox is a clear example of an unclear truth; we can in fact prove that the liar is unclear from the theory we designated BCT. The liar paradox, therefore, does not witness the unclarity of the distinction between clear and unclear truth. Here we shall argue that the revenge sentence R='R is not clearly true' does.

The question at stake is whether it is always a clear matter whether p is clear or not. This can be broken down into two claims; that when it is clear that p, this fact is itself clear, i.e.

4 If it's clear that p then it's clear that it's clear that p.

and that when it's unclear that p this fact is itself clear, i.e.

5 If it's not clear that p then it's clear that it's not clear that p.

In general the distinction between clear F's and unclear F's is not a clear one. This phenomenon manifests itself quite evidently in the paradox of the heap. When presented with a Sorites sequence for the property of being a heap, i.e. a sequence of piles of sand beginning with a few grains and ending with a large mound, we may similarly introduce the distinction between the piles which are clearly heaps and those which are not. The distinction is evidently there, as the enormous pile at the end of the sequence is not only a heap, but a clear heap, whereas the first pile is clearly not a heap, and therefore not a clear heap. But when do the clear heaps stop being clear heaps? This distinction, between the clear heaps and the unclear heaps, is no clearer than the distinction between heaps and non-heaps. At some point there must be a clear heap which is not clearly a clear heap, or an unclear heap which is not clearly an unclear heap; either we have a failure of 4 or 5 (and most probably both.)

The situation is somewhat similar in the context of the revenge sentence. In the last section we noted that the standard revenge argument begins by assuming that it was clear that R is not clearly true and deriving a contradiction. We then observed that we could not in fact prove a contradiction, we could only show that R is true but not clearly true. It is important to notice that the conclusion of this argument can never be clear; there are never clear examples of things which are true but not clearly true. This is a result of a general argument known as Fitch's paradox: if it were clear that R was true but not clearly true, then, since a conjunction can be clear only if both conjuncts are, it would be clear that R was true contradicting the clarity and therefore truth of the second conjunct. However in BCT, if the argument from p to q is valid, then so is the argument from it's clear that p to it's clear that q. Since the conclusion of the argument cannot be clear, neither can the premise.

The upshot of all this: if r is chosen so that we can show the sentence 'r if and only if 'r' is not clearly true' (this can be achieved by diagonalisation or setting 'r' to 'R is not clearly true') then one can prove in BCT

It's not clearly clear that r.

The revenge paradoxes, in this setting, do not result in inconsistency, but second order unclarity.

It's natural to wonder what happens if we consider stronger liars obtained by iterating the number of clearly operators. Let us write 'clearly^n' to indicate n 'clearly's in a row. Suppose that, by diagonalisation, we chose a sentence r_n so that we can prove 'r_n if and only if it's not clear^n that 'r_n' is true'. Then we can show:

It's not clearly clear^n that r_n

In other words, the nth revenge sentence gives rise no $n+1$th indeterminacy in BCT. We shall come back to these questions in a more rigorous setting in section 7.3.2.

7.2.3 Assertion

This brings us to another benefit this approach has over its rivals. As we noted earlier one of the forms a revenge paradox can take involves the propositional attitudes of acceptance and rejection, or, sometimes, permissible assertion and denial. The problems are often discussed in the context of non-classical solutions (see Soames [125], Field [40], Beall [8], Priest [102].), but they seem to be just as bad for classical solutions (see especially Feferman [34], Maudlin [89].)[14] For example, paracomplete logicians – logicians who relinquish the law of excluded middle – are at pains not to accept (or assert) that the liar is neither true nor untrue since, by the deMorgan laws, this would commit them to accept that the liar is both true and untrue. In order to express their disavowal of LEM they *reject* (deny) the claim that the liar is true or untrue. Rejecting (denying) p is not to be reduced to acceptance (assertion) of the negation of p. These theorists need then to say something about the sentence: A='A is not assertable.'

How might paradoxes involving assertion and acceptance arise in this setting? A similarly central distinction between assertable (acceptable) propositions exists in the present framework: like the paracomplete logician we agree that one should not assert that the liar is true, or that it's untrue since it is unclear. In general, we might endorse the following schema

R. If it's unclear whether p then you should not assert that p and you should not assert that $\neg p$.

Perhaps this is a basic fact about unclarity. However it is natural to think that it could be explained by the fact that unclarity prevents knowledge. We should not assert that the liar is true or assert that the liar is untrue because we do not, and cannot, know which it is.

Note that R is an externalist norm in the sense that one is not always in a position to know whether one is complying with it. This can happen for quite mundane reasons – for example if you do not know whether Harry is clearly bald because you simply do not know how much hair he has. In these cases someone else in a better epistemic position than you would be able to evaluate your assertion for compliance.[15] However we must also make room for the possibility that sometimes no-one else is, or even could be, in a position to evaluate whether you have complied with R or not; for example if it is unclear whether p is unclear or not.

How might assertability paradoxes arise in this framework? Note first that the problematic notion is a normative one, not a descriptive one. The sentence A = 'it's not the case that so-and-so has at some point asserted that A is true' is clearly true or false depending on what the person in question has asserted.[16] The problem supposedly arises when we want to talk about which propositions I should and shouldn't assert. Now evidently there are cases where I shouldn't assert p even though it's clear that p; I shouldn't assert that there are an even number of stars within a hundred light years of the sun, nor should I assert that there are an odd number, since I do not know either way, even though it is a clear matter whether or not there are an even number of stars within a hundred light years of the sun. However the proposition that the liar isn't true seems different; it seems to be *inherently* unassertable. The reason that it is unassertable is to do with unclarity about whether the liar is true or not.

Is there a problem of expressing the distinction between inherently unassertable propositions in a semantically closed language without falling afoul of assertability paradoxes? A proposition is inherently unassertable when the source of its unassertability is the kind of unclarity we find generated by semantic paradoxes; the problematic notion of unassertability is directly controlled by the clarity operator. Thus we get a partial converse of R

[14]The problem of assertion and acceptance does not apply exclusively to philosophers in this list.

[15]Usually an assertion such as this will fail to comply with the norm of asserting only what you know.

[16]Paradoxes formulated using the operator 'So-and-so is disposed to assert that p' look like they are equally unproblematic.

(*) p is inherently unassertable if and only if it's unclear whether p.

It seems that, as long as we can consistently introduce the clarity operator into the language, we can also introduce the notion of inherent unassertability into the language. And furthermore, it seems that the assertability paradox A='it is inherently impermissible to assert that A is true' should pattern with the revenge paradox R='it is not clear that R is true'.

Let p be a proposition such that it is unclear whether p is clear or not (for example, the proposition that R is not clearly true). Assuming that (*) is determinate, we can infer that if it's unclear whether it's unclear whether p, then it's unclear whether p is inherently unassertable or not. This follows from the fact, provable in BCT, that if one side of a clear biconditional is unclear, the other side is unclear.

Intuitively there are a bundle of tightly related phenomena which arise in the cases of unclarity. When p is such a case there is a peculiar source of ignorance and uncertainty, it is impermissible to assert or question whether p, and so on; we introduce the word 'unclear' to distinguish propositions that associate with that bundle of properties. What happens when it's unclear whether p is clear or not? Since 'unclear' was introduced as a way of picking out propositions with that bundle of properties, to say it's unclear whether p is clear or not is to say that it's unclear whether p has that bundle of properties. In particular:

1. If it's unclear whether p is clear or not, then it's unclear whether it's possible to know whether p.

2. If it's unclear whether p is clear or not then, provided you are as knowledgeable as you could be, it's unclear whether it's permissible to assert p or not.

7.2.4 Can we recover sentential clarity?

Suppose I look into the night sky at the Orion constellation and see that Orion's belt consists of three stars. It should be clear, I hope, that in doing this I do not see, or stand in any obvious kind of relation to, a *sentence*. One can see that Orion's belt has three stars without having to speak any particular language, which makes it hard to imagine how seeing that Orion's belt has three stars could be constituted by any relation between me and a sentence. 'Sees that' belongs to a class of operator expressions, including also 'hopes that', 'desires that', 'says that', 'fears that', 'knows that', even 'it's not the case that', which are not easily reducible to, and whose work could not be done by a predicate of sentences.

I have been suggesting that the word 'clearly', as introduced in the last section by its role in classifying the paradoxes, belongs to this class. An interlocutor might object that the use of a clarity operator commits us to the coherence of a corresponding 'linguistic clarity predicate' which obeys paradox generating principles. Certainly there are a number of sentential predicates you can define, once you have a clarity operator, but, I shall argue, they cannot play the same role as the clarity operator and furthermore that they are paradox generating only if we assume the naïve theory of truth and saying encapsulated by disquotational schemata S and T which are already known to generate paradoxes on their own.

I shall begin by considering whether there could be is a sentential predicate which plays the same role as the clarity operator. We'll use the terminology:

Def. s is definite as currently used in English.

Saying what it means for a predicate to 'play the same role as' an operator is a delicate matter. However I take it that it should at least perform the same classificatory role as the clarity operator.

However it seems that any predicate of the form above cannot play this role. For example, let T be any sentence that means in German that T is true in German. For my

purposes it does not matter what T is, however I believe that setting T equal to the German sentence 'T ist in Deutsch wahr' would do. T is a truth-teller for German, much like the sentence $T' =$ 'T' is true in English' is a truth-teller for English.

Is T true in German or not? The answer, one would have thought, is that it's unclear in the same way that it's unclear whether T' is true in English. Note that in classifying these two cases as alike we relied on the fact that what is unclear in each case is not a sentence of English or German, but the proposition that that sentence is true in English, or German, respectively. The adverb 'clearly' can modify any predicate, for example 'white' in the earlier example, but also 'true in English' and 'true in German'. Thus we may say:

> T is not clearly true in German.
>
> T' is not clearly true in English.

The thing that is unclear is whether T is true in German which, unlike the German sentence T, is the kind of thing that can be said, rather than the kind of thing that is used to say things. It is not easy to see how this important parallel between the German and English truth-teller could be made using the predicate 's is definite in English.' The claim that T is indefinite in English, the obvious parallel to the claim that T' is indefinite in English, will not do: T doesn't refer to a sentence of English so it is not indefinite in English.[17]

Other aspects of the role of clarity are also hard to capture with a predicate. For example, many philosophers have suggested that it is not possible to know whether p if it's unclear whether p. My own positive theory, outlined in more detail in the next section, situates the notion of clarity within a theory of rational propositional attitudes, which partly includes the role that beliefs about the unclarity of p play in regulating your beliefs and desires about p. It is natural to state this theory using an operator. Any theory that states a similar regulation of propositional attitudes using a sentential predicate must postulate implausible connections between what can be rationally known, believed and desired and your beliefs about the the way particular public language sentences are used.

Even though there is a good case to be made that the most perspicuous way to model the revenge paradoxes is to use an operator, many readers will no doubt be feel that, perspicuous or not, there must be *some* way to paraphrase 'it's clear that p' using a predicate expression taking a sentence as an argument. For example, Prior points out that 'although *breaking one's leg* is quite obviously not a relation between a person and any form of words whatever', in particular it is not a relation between me and the expression 'my leg', one certainly stands in *some* relation to this expression when one breaks one's leg. For example, I have broken what would in this context be denoted in English by the expression 'my leg'.

If there were a sentential predicate, $Def(\ulcorner\phi\urcorner)$, which was everywhere intersubstitutable with 'it's clear that ϕ', we would be susceptible to the various paradoxes including Montague's paradox and theorem 7.1.4 that we outlined for sentential predicates in section 7.1.

One must be careful to distinguish the different senses of 'paraphrase.' In general one might think that one could somehow introduce a predicate into English (or whatever our language happens to be), 'is definite in English', such that for each proposition p there is a sentence s such that the claim that s is definite in English is equivalent to the claim that it's clear that p. I do not think that this principle is problematic for the revenge type reasons we have been discussing (although it is probably false because there are more propositions than English sentences.) We need a stronger form of paraphrase if we are to be able to prove an instance of the SRT. schema from the RT. schema – we need that the claim that it's clear that p to be equivalent to the claim that a *particular* English sentence, 'p', is

[17] We could break the parallel between the English and German truth-teller by instead saying 'T is true in German' is indefinite in English, or by saying T is indefinite in German, but these are subject to other problems. For example the former would not do since T would not be clearly true in German, even if English speakers used 'T is true in German' in such a way as to make it definite – perhaps by using it to mean that snow is white.

definite in English. Here p ranges over sentences in the *expanded* language containing both the definitely predicate and clarity operator.

We shall consider some putative ways of achieving such a paraphrase in what follows. However, it is worth starting by considering a simpler more familiar operator that we know cannot be paraphrased with a sentential predicate: negation. A general theorem establishes that there is simply no sentential predicate which paraphrases, in the strong sense, an operator that satisfies the principles below.

$(\phi \rightarrow \blacksquare\phi) \rightarrow \blacksquare\phi$

$(\blacksquare\phi \rightarrow \phi) \rightarrow \phi$

$\phi \rightarrow (\blacksquare\phi \rightarrow \psi)$

For suppose $F(\ulcorner\phi\urcorner)$ was equivalent to $\blacksquare\phi$. By diagonalisation we have $\gamma \leftrightarrow F(\ulcorner\gamma\urcorner)$ (*). Then we have $F(\ulcorner\gamma\urcorner)$ by the first principle and the paraphrase of the left to right direction of (*), and γ by the second principle and the paraphrase of the right to left direction of (*). Thus, by the last principle we have ψ for arbitrary ψ. Since negation satisfies all these principles there is no predicate which paraphrases it in the strong sense defined above.[18]

In general it is not possible to simply introduce a falsity predicate, F, whose logic mirrors the negation operator; nor, I claim, is it possible to introduce a definiteness predicate, Def, whose logic mirrors the clarity operator. It is worth reflecting on why two of the most obvious strategies for introducing a predicate that is equivalent to the clarity or the negation operator fail. The first strategy is a version of Prior's paraphrase of 'I broke my leg' – the predicate 's says something in English which is clear (or not the case)' seems to approximate the relevant operator expressions. We might therefore hope to define:

> 'ϕ' is false in English iff for some p 'ϕ' means that p in English and it's not the case that p
>
> 'ϕ' is definite in English iff for some p 'ϕ' means that p in English and it's clear that p

The second strategy uses the truth predicate

> 'ϕ' is false in English if and only if it's not the case that 'ϕ' is true in English.
>
> 'ϕ' is definite in English if and only if it's clear that 'ϕ' is true in English.

If our definitions of sentential falsity and definiteness really do correspond to negation and clarity we should be able to show

> 'ϕ' is false in English if and only if it's not the case that ϕ
>
> 'ϕ' is definite in English if and only if it's clear that ϕ

It should be noted that even if we could prove these principles from contingent facts about the truth and meanings of sentences in English, they would at best be contingent equivalences. For example, had we used 'snow is blue' to mean that snow is white then 'snow is blue' would not be false in English, according to either paraphrase, even though it's not the case that snow is blue.

That said, one might be able to prove these instances from contingent facts about English. Let us make the following assumptions

1. 'snow is white' means that snow is white in English

[18]Similar points can be made about necessity operators and knowledge operators (see Montague [95], Montague and Kaplan [68]) but I think the case is strongest for negation (see Deutsch [22].)

2. for all p, q if 'snow is white' means that p and 'snow is white' means that q then p if and only if q (indeed $p = q$.)

It is then simple to prove that according to the first paraphrase, 'snow is white' is false in English if and only if it's not the case that snow is white. Similarly if we were to assume the contingent fact that

3. 'snow is white' is true in English if and only if snow is white.

we could prove that according to the second paraphrase 'snow is white' is false in English if and only if it's not the case that snow is white. Parallel arguments could be made for the material equivalence of the claim that 'snow is white' is definite in English and the claim that it's clear that snow is white relative to each definition of 'definite'.

Both of these derivations, however, rely on particular facts about the meaning of sentences in English, meanings which happen to be disquotational. However the full disquotational schema, T, of which 3. is an instance is known to be inconsistent, as is the combination of 2 with the general disquotational schema, S, which 1 is an instance of. It should be obvious, therefore, why the equivalences do not hold in full generality; the instances where the equivalence fail are *exactly* the instances in which the disquotational principles of naïve semantics fail. Saying that 'ϕ' means something clear in English is not the same as saying that it's clear that ϕ. They amount to the same thing only when 'ϕ' means that ϕ. Saying that 'ϕ' is clearly true in English amounts to saying that it's clear that ϕ only when ϕ and 'ϕ' is true in English are intersubstitutable.

7.3 A theory of clarity

In the previous section I introduced the word 'clearly' in a non-partisan way, intending the discussion to be neutral between the many different interpretations we could substitute for 'clearly.' While everything I said there would be compatible with a purely epistemic reading of 'clearly', for example, I shall spend this section developing my own positive theory of clarity.

The kind of phenomena that we are interested in includes not just paradoxicality, characterised by failures of the various disquotational schemata, but the more general kind of phenomenon associated with non-paradoxical, but somehow factually defective discourse involving truth. Consider for example:

L_1 L_2 is not true in English

L_2 L_1 is not true in English

While there is no inconsistency involved here, even if we assume the disquotational schema, something is not quite right. There seems to be no plausible semantic constraints, including the relevant instances of T, which would decide whether or not L_1 or L_2 is true or not. Furthermore the non-semantic facts do not help decide whether they're true or not either. I suggest that there is no fact of the matter whether L_1 or L_2 are true or not.

Once we have seen that some discourse involving truth is factually defective it is natural to ask whether talk involving paradoxical liar-like sentences can be non-factual too. It is perhaps easier to evaluate whether cases like the truth-teller and L_1 and L_2 involve non-factuality since there are multiple truth values we could assign that are consistent with any plausible semantic constraint, including the relevant instances of the T-schema. Once we have given up the T-schema, however, isn't any truth-value assignment game? I think this would be too quick: the paradoxes provide no reason to give up on the constraint that possible truth-value assignments obey the normal clauses for the extensional connectives. These constraints would guarantee that the liar was either true or false and not both, but they would not tell us which, inviting the same thoughts we had about the truth-teller. It's

also worth noting that the liar sentence, L, is just a special case of the above example in which $L_1 = L_2$. I think it is natural to think that the truth value of the liar sentence is also unconstrained. Just like the truth-teller it seems unclear what kind of evidence would settle the question of whether or not the liar is true, it seems like we shouldn't assert that it's true or that it's false, and there generally doesn't seem to be any fact of the matter concerning its truth or falsity.

The kind of predicament we find ourselves in when we have all the evidence we could have about a subject matter and we still seem unable to decide whether p is often associated with cases in which p is borderline. A natural conjecture is that there is a general phenomenon, indeterminacy concerning that subject matter, which encompasses both the kinds of situations we find ourselves when confronted with questions about whether T true in German or whether L_1 is true in English, and so on, and with questions about whether a borderline bald man is bald or not. A characteristic feature of cases of indeterminacy is that they involve an inability to know whether p even when all the facts are available to you. Furthermore there is a certain kind of irrationality involved when someone cares intrinsically about the indeterminate. Facts such as these situate indeterminacy within a theory of rational propositional attitudes. The source of my ignorance concerning whether T is true in German, my inability to rationally care, believe, and so on, that T is true in German are not facts about the German or the English language – I need not be acquainted with either language. Indeterminacy therefore satisfies one of the deciderata for providing an interpretation of clarity, as introduced informally in section 7.2.1: it is a theory of *propositions* and provides us with a way of assessing which propositions are suitable truth conditions to put in the disquotational schema.

This way of thinking about things is perhaps more familiar in the philosophy of vagueness. While the majority of theories of vagueness appear to be linguistic theories - examples falling under this this camp include semantic indecision theories (e.g. McGee McLaughlin [92]), metalinguistic safety accounts (Williamson [138]) and inconsistency theories (Eklund [29]) - there is precedent for the operator view in Field [37], Fine [43] 1975, Barnett [4].

In chapter 5 provided an explicit definition of determinacy. We do not actually need to be this explicit: we can introduce this operator implicitly to someone who doesn't understand it by outlining its inferential properties and the role it plays in a theory of rational propositional attitudes:

> **Logical** The operator 'it's determinate that p' is governed by the modal logic KT, or some extension of it:
>
> > **Closure** If it's determinate that if p then q, and it's determinate that p, then it's determinate that q.
> >
> > **Factivity** If it's determinate that p, then p
> >
> > **Necessitation** If one can prove p from classical logic and these three principles, one can prove that it's determinate that p.
>
> **Alethic** Determinacy licenses disquotational reasoning. If it's determinate whether p, then 'p' is true just in case p.
>
> **Epistemic** If it's indeterminate whether p, then it's not rationally known whether p.
>
> **Doxastic** No rational person who isn't certain that it's not indeterminate whether p can rationally assign a credence of 1 or 0 to p.
>
> **Pragmatic**
>
> > **Assertion** It is not permissible to assert that p (that $\neg p$) if one believes that it's indeterminate whether p.

Questions It is not permissible to as whether p if you have been told that it's indeterminate whether p.

Bouletic It is not rational to care intrinsically about the indeterminate: if the strongest determinate proposition p entails is identical to the strongest determinate proposition q entails then you should be indifferent between p and q, for maximally specific propositions p and q.

Attitudes It is not rational to wonder whether, hope that, fear that, [...] p, if you believe that it is indeterminate that p.

We called this the determinacy role.

The role described above is not the only reasonable way to go about developing a theory of indeterminacy. In [37] Hartry Field argues for alternative constraints on our doxastic attitudes. He argues among other things that:

Reject If you know that it's indeterminate whether p you should reject p and reject $\neg p$.

Credences Your credence in p should be the same as your credence in Δp. In particular if $Cr(\nabla p) = 1$ then $Cr(p) = Cr(\neg p) = 0$.

The role I have described states that when you are certain that p is indeterminate you should have an intermediate credence, where as according to Field we should have a credence of 0 in both p and its negation. Field's proposal would also fit well with the theory I am sketching.

7.3.1 The theory DFS

The determinacy role stipulates that determinacy interacts with truth in a certain way: it requires that it obey the basic theory of clarity and truth, BCT, we mentioned in section 7.2.1. However our remarks in section 7.3 motivated a much stronger compositional theory of truth and determinacy which we shall now formulate.

Let $\mathcal{L}^-$ be the language of arithmetic with a function symbol for each primitive recursive function f. Let $\mathcal{L}$ be $\mathcal{L}^-$ augmented with a truth predicate, Tr, and primitive determinacy operator, Δ. We shall retain the conventions we set in place in section 7.1 regarding Gödel numbering and the dot notation. The numeral for Gödel number of a formula ϕ is written $\ulcorner\phi\urcorner$, and there are primitive recursive functions $\dot{\neg}$ ($\dot{\wedge}, \dot{\vee}$ and so on) taking the Gödel number of a formula to the Gödel number of the result of prefixing negation to that formula (and similarly for $\dot{\wedge}, \dot{\vee}$ etc.) In arithmetic we can define predicates $Sent$, At and Ver to saying that a number is the Gödel number of a sentence, atomic arithmetical sentence and an atomic arithmetical truth respectively. With this in place we can state our theory, which we call DFS

PA. Peano arithmetic including full induction (i.e. the induction schema may take formulae containing the truth predicate and the determinacy operator.)

ΔNec If $\vdash \phi$ then $\vdash \Delta\phi$

K $\Delta(\phi \to \psi) \to (\Delta\phi \to \Delta\psi)$

T $\Delta\phi \to \phi$

BF $\forall x \Delta\phi \to \Delta\forall x\phi$

RT $(\Delta\phi \vee \Delta\neg\phi) \to (Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi)$ when ϕ is closed.

At. $\forall x(At(x) \to (Tr(x) \leftrightarrow Ver(x)))$

C→. $\forall x \forall y(Sent(x) \wedge Sent(y) \rightarrow (Tr(x \dot{\rightarrow} y) \leftrightarrow (Tr(x) \rightarrow Tr(y)))$

C∨. $\forall x \forall y(Sent(x) \wedge Sent(y) \rightarrow (Tr(x \dot{\vee} y) \leftrightarrow Tr(x) \vee Tr(y)))$

C∧. $\forall x \forall y(Sent(x) \wedge Sent(y) \rightarrow (Tr(x \dot{\wedge} y) \leftrightarrow Tr(x) \wedge Tr(y)))$

C∀. $\forall x(Sent(x(\bar{0}/v)) \rightarrow (Tr(\dot{\forall} vx) \leftrightarrow \forall y Tr(x[\dot{y}/v])))$

C¬. $\forall x(Sent(x) \rightarrow (Tr(\dot{\neg} x) \leftrightarrow \neg Tr(x)))$

CΔ. $\forall x(Sent(x) \rightarrow (Tr(\dot{\Delta} x) \leftrightarrow \Delta Tr(x)))$

Nec If $\vdash \phi$ then $\vdash Tr(\ulcorner \phi \urcorner)$[19]

Conec If $\vdash Tr(\ulcorner \phi \urcorner)$ then $\vdash \phi$

Let me begin by citing an important result (for the proof see the appendix.)

Theorem 7.3.1. *The system* DFS *is consistent.*

In fact one can go further and show that it doesn't prove any false arithmetical sentences. If you restrict induction to its arithmetical instances it proves no new arithmetical theorems.[20] It's also consistently augmentable with propositional quantification (and full second order logic.)

DFS has as its truth theoretic basis the compositional theory of truth known as FS (see Halbach [55]) which consistents of the principles not mentioning Δ – i.e. the compositionality axioms apart from CΔ (i.e. those beginning with a 'C'), At, and Nec and Conec. Compositionality is a fundamental principle of modern semantics and it is a significant drawback of many rival theories that they do not have it. A noteworthy consequence of compositionality, i.e. of the C-axioms, is the principle of bivalence: every sentence is either true or false (a sentence is false if it has a true negation.)[21] Note also that compositionality extends also to the intensional connective Δ (this will prove useful when we augment the language to state its own semantic theory.)

We have already motivated the principles ΔNec, K, T and RT in our discussion of BCT (although I have more to say about ΔNec.) RT tells us when we can and can't enquote and disquote. In any modal logic based on classical quantification theory the converse of BF, CBF, will already be a theorem, as will be the claim that there is no indeterminate existence: $\Delta \forall x \Delta \exists y x = y$. Although I have no knock down argument for further including BF in the theory, it seems natural, and since it is consistent there is no immediately compelling reason to exclude it.

The unrestricted principles of necessitation and conecessitation Nec, Conec and ΔNec are much more tricky. The principle ΔNec guarantees that if ϕ is a theorem of the theory then so is $\Delta\phi$, $\Delta\Delta\phi$, and so on. One might think this is far too strong; the theory contains contingent semantic principles which depend on the way we use the symbols ¬, ∨ and so on. A further problem is that these principles are in fact ω-inconsistent. Firstly note that already the truth predicate satisfies the conditions for McGee's paradox 7.1.3. However the theory contains a distinct ω-inconsistency that involves only principles involving the determinacy operator:

Theorem 7.3.2. BF + T + ΔNec + *RT* + *CΔ., are ω-inconsistent in PA.*

[19] The rule Nec is actually redundant; it is a derived rule of BCT.

[20] These facts follow from the results of Halbach [55]. One can interpret DFS into the system FS Halbach considers by defining $\Delta\phi := \phi \wedge Tr(\ulcorner \phi \urcorner)$.

[21] This follows from C∨, C¬ and the truth of the principle of excluded middle, which we get by applying Nec to the principle of excluded middle.

To prove this we consider the instance of the diagonal lemma: $\delta \leftrightarrow \neg\forall n \Delta Tr(\ulcorner \Delta^n \delta \urcorner)$ – I shall not go into the proof however. Here I briefly suggest a number of ways around this problem. My preference is to restrict the necessitation principles, but I shall not spend much time debating the various pros and cons.

Consider the rule

KTNec If ϕ is provable in KT (i.e. from T, K and KTNec) then infer $\Delta\phi$

KTNec allows us only to necessitate the theorems of classical logic and KT. Let DFS_n be the system DFS where all the necessitation principles are restricted to at most n applications in a proof, plus KTNec which allows arbitrarily many applications if the sentence is provable without using the truth principles.

We also define three subsystems of DFS:

S_1. DFS - BF - Conec - $C\forall$.

S_2. DFS - $C\Delta$. - Conec - $C\forall$.

S_3. DFS - $C\Delta$. - BF - $C\forall$.

then

Theorem 7.3.3. *The theories* DFS_n*,* S_1*,* S_2 *and* S_3 *have standard models*

The proofs are quite technically involved so I shall not present them here.

7.3.2 Higher order indeterminacy and revenge

Note that we can formally prove many of the facts indicated in our earlier discussion. In particular, if via standard diagonalisation techniques we generate sentences $\lambda, \gamma, \gamma_n$ satisfying: $\lambda \leftrightarrow \neg Tr(\ulcorner \lambda \urcorner)$, $\gamma \leftrightarrow \neg\Delta Tr(\ulcorner \gamma \urcorner)$ and $\gamma_n \leftrightarrow \neg\Delta^n Tr(\ulcorner \gamma_n \urcorner)$ then we can prove

1. $\nabla\lambda$
2. $\neg\Delta\Delta\gamma$
3. $\neg\Delta\Delta^n\gamma_n$

Here ∇p is defined as $\neg\Delta p \wedge \neg\Delta\neg p$ and $\Delta^n p$ is p prefixed by n Δ's. These results agree with Field's recent diagnosis of the revenge liars and their strengthenings [40].

This raises some interesting questions about higher order indeterminacy in this theory. One can rule out various kinds of higher order indeterminacy by adding a number of principles including:

4 $\Delta\phi \rightarrow \Delta\Delta\phi$

5 $\neg\Delta\phi \rightarrow \Delta\neg\Delta\phi$

B $\phi \rightarrow \Delta\neg\Delta\neg\phi$

Perhaps unsurprisingly, none of these principles can be consistently added to DFS.

Theorem 7.3.4. DFS *+ 4 is inconsistent.*

Proof. 1. $\gamma \leftrightarrow \neg Tr(\ulcorner \Delta\gamma \urcorner)$ instance of the diagonal lemma.

2. $\Delta\gamma \rightarrow \Delta\Delta\gamma$ by 4.
3. $\Delta\Delta\gamma \rightarrow Tr(\ulcorner \Delta\gamma \urcorner)$ by RT.
4. $\Delta\gamma \rightarrow Tr(\ulcorner \Delta\gamma \urcorner)$ 2 and 3.

5. $\Delta\gamma \rightarrow \neg Tr(\ulcorner \Delta\gamma \urcorner)$ by T and 1
6. $\neg\Delta\gamma$ by 4 and 5.
7. $Tr(\ulcorner \neg\Delta\gamma \urcorner)$ by necessitation of Tr.
8. $\neg Tr(\ulcorner \Delta\gamma \urcorner)$ by truth functionality of $\neg$.
9. γ by 1
10. $\Delta\gamma$ by necessitation for Δ, which contradicts 6.

□

Theorem 7.3.5. DFS *+* B *is inconsistent.*

Proof. 1. $\gamma \leftrightarrow Tr(\ulcorner \Delta\neg\gamma \urcorner)$ instance of the diagonal lemma.

2. $\gamma \rightarrow \Delta\neg\Delta\neg\gamma$ by B
3. $\gamma \rightarrow Tr(\ulcorner \neg\Delta\neg\gamma \urcorner)$ by 2 and RT.
4. $\gamma \rightarrow \neg Tr(\ulcorner \Delta\neg\gamma \urcorner)$ by truth functionality of $\neg$.
5. $\gamma \rightarrow Tr(\ulcorner \Delta\neg\gamma \urcorner)$ by 1.
6. $\neg\gamma$ by 4 and 5.
7. $\Delta\neg\gamma$ by necessitation for Δ
8. $Tr(\ulcorner \Delta\neg\gamma \urcorner)$ by necessitation for Tr
9. γ by 1, which contradicts 6.

Since DFS+$\mathsf{5}$ entails B it follows that we cannot add $\mathsf{5}$ to DFS either. □

7.3.3 Hierarchies of determinacy operators

In his recent critical notice of Hartry Field's 'Saving Truth from Paradox', Priest raises a number of problems with Field's project, and with his hierarchy of determinacy operators.

Priest argues that Field cannot express the general notion of defectiveness that the liar and its relatives have. Since this is an important way of misunderstanding my own view, and marks a significant difference between my own view and Field's I shall quote him at length:

> "There are two jobs for the notion of defectiveness to do: [a] we must be able to say of certain sentences, e.g., the liar, that they are of this kind; and [b] we must be able to talk about such sentences in general and say things about them. [...].
>
> Now, Fields D operator does the job of [a] in many cases. As we have seen, for a number of defective sentences, A, like the Liar, we can say truly $\neg DA \wedge \neg D\neg A$, or at least $\neg D^{\alpha}A \wedge \neg D^{\alpha}\neg A$, for some suitable iteration of 'D's maybe into the transfinite. But the D operator cannot do the job of [b]: as α increases, the extension of $\neg D^{\alpha}$ gets larger and larger (p. 238), so for no α does the extension of $\neg D^{\alpha}$ comprise all the non-(determinately true) sentences. Where $Q = \neg D^{\alpha}Tr(\ulcorner Q \urcorner)$, Q is not in the extension of $\neg D^{\alpha}$. Nor is it possible to define a predicate whose intuitive meaning is something like $\bigvee\{\neg D^{\alpha}Tr(x) \mid \alpha$ is an ordinal$\}$, since, as Field shows, the precise definition of this depends on some ordinal notation, and will therefore take us only so far up the ordinals. [...].

> Indeed, without the notion, one cannot even formulate the driving thought behind Fields own solution: that the LEM fails because of the existence of indeterminacies. To declare all general claims about indeterminacy unintelligible is an act of ladder-kicking-away desperation of Tractarian proportions." Priest, 'Hopes Fade for Saving Truth' [104]

Field's picture as Priest describes it is that there is a (large) hierarchy of predicates, and each level only does a partial job of characterising the true notion of defectiveness or being diseased which includes the liar, the revenge liar, and the further iterations.

But what reason do we have to think that, say, the revenge liar is defective?[22] A poor reason would be to say that it's also self-referential: we know that self-reference is neither necessary nor sufficient for defectiveness.

This is problematic for Field since it suggests that there is a notion – which both Field and Priest informally call 'defectiveness' – which Field cannot express. A sentence is defective when it falls under one of Field's partial predicates; the union of all Field's partial defectiveness predicates therefore seems to be the notion that is being informally employed, but which is not expressible within Field's framework.

The view that I am defending does not appeal to any partial notion of defectiveness; the notion they are calling defectiveness can be completely expressed using the indeterminacy operator. I therefore reject Priest's assumption that the revenge liar is defective – i.e. that it's indeterminate whether R is true, where R='R is not determinately true'. I reject this because this claim is *itself* indeterminate; asserting this would be akin to asserting that someone is bald when they have a borderline number hairs. I have argued above that if a sentence is indeterminate one should not assert either it or its negation. Furthermore, I have shown that it's an indeterminate matter whether the revenge liar is determinately true or not. So, *pace* Priest and Field I would object to the claim that the revenge liar is indeterminate in exactly the same way that I would object if someone asserted that the liar is true. According to the analogous objection leveled at my view, Priest is asserting the unassertable and reasoning from that in a way that is impermissible.

The relevant notion of indeterminacy, the one I've axiomatized above and which governs assertability, knowability and so on, is perfectly well captured by the operator ∇. Like many useful but vague concepts it admits indeterminate instances. It would be a mistake, however, to assert that these indeterminate cases of indeterminacy are straightforwardly indeterminate. That would be to assert the indeterminate - to assert of anything that is indeterminately F that it is F is to assert the indeterminate, and it would be equivocation to use the word 'indeterminacy' when one really means 'indeterminacy at some order'. The equivocation is plain for all to see when in the analogous objection levelled at the present view: 'as α increases, the extension of $\neg D^{\alpha}$ gets larger and larger, so for no α does the extension of $\neg D^{\alpha}$ comprise all the non-(determinately true) sentences.' This would not make sense if we were using the operator 'D' and the word 'defective' interchangeably, for when $\alpha = 1$, $\neg D^{\alpha}Tr(\cdot) = \neg DTr(\ulcorner\phi\urcorner)$ comprises all the defective sentences simply by the fact that D was introduced to mean 'defective'.

Note that if we did interpret Priest's use of the word 'defective' to just *mean* indeterminate at some order or other, his final remarks no longer hold any bite. It may well be the case that nothing is determinate at all orders (this kind of view naturally corresponds to the theories DFS_n which I describe in section 7.3.1.)

Given that I have been denying that the higher iterations of the indeterminacy operator play an important role in the workings of the liar paradox, and in particular, denying that they the serve the purposes required by a solution to the semantic paradoxes, it would be a fair to ask in which respects the second order, third order, ..., nth order indeterminacy differ.

[22] Remembering that 'defective' or 'indeterminate' is the notion I've so far been characterising by the Δ-role, i.e. its inferential and normative properties, and is not necessarily a notion governed by intuitions Priest is appealing to.

It is easy to verify that if on operator O satisfies the axioms of the theory DFS, then so will the iterations of this operator: OO, OOO, and so on. This naturally raises the question of what the intended interpretation of 'Δ' is, and if there is anything we can say to narrow down this infinite list of candidate interpretations. One way to do this is to appeal to the role that Δ plays in a theory of propositional attitudes; in this respect I have suggested that it satisfies the Δ-role. As it is currently stated, the Δ-role consists of several conditionals stating that indeterminacy poses a distinctive obstacle to knowledge, that one cannot rationally believe or care about a proposition you believe to be indeterminate and so on. While iterations could satisfy these conditionals as well, it is natural to want to introduce indeterminacy as a name for *whatever* that distinctive obstacle to knowledge is, or for whatever it is that regulates your beliefs in such and such a way, and so on. Thus we should strengthen some of the conditionals in the Δ-role to *biconditionals*. If the distinctive barrier to knowing whether p characteristic of cases like the liar is present in p then p is indeterminate. This is what distinguishes the determinacy operator from its iterations.

We can contrast this view with an alternative position. According to the alternative there are infinitely many operators that satisfy the Δ-role: if O is an operator that satisfies the Δ-role so does all of its iterations. This appears to be Field's position in a number of writings (see, for example [39].) As stated the present view maintains that there is only one true notion of defectiveness at play here, and the others are just iterations of this notion and that this can be enforced by strengthening some of the conditionals prescribed by the Δ-role to biconditionals. The alternative position asserts that second order indeterminacy is also a distinctive source of ignorance, and would therefore want only half of the biconditional: it is possible to know that p if and only if it's determinate that p, since p might be determinate but not knowable because it is not determinately determinate. However the idea that indeterminacy is not the only distinctive obstacle to knowledge invites the thought that whatever the obstacle is in all these cases we could surely introduce an expression for it (this is, effectively, Priest's intuition.) And if so, why not just use a more expansive use of the word 'indeterminate' or 'defective' from the beginning? Field's position that that there are all these distinct kinds of obstacles to knowledge, assertion, rational belief and so on, but there's no umbrella phrase for them (on pain of paradox) is reminiscent of the Tarskian strategy of denying the expressibility of seemingly intelligible notions. The alternative suggestion, which seems to be extremely natural, is that there is an umbrella expression for this, although it's one that does not iterate trivially.

7.3.4 Assertion

With this in mind, let us consider an example. Given that it's indeterminate whether it's determinate that p, may we assert p, or should we abstain from assertion? A naïve but persuasive thought is that second order indeterminacy in p indicates a *kind* of badness in p and one should accordingly refrain from asserting p or its negation in all contexts just as one should refrain in cases like the liar and the truth-teller. This thought, however, is not sustainable. Second order indeterminacy indicates indeterminacy over whether p is "bad" in the relevant sense, and thus, indeterminacy over whether one should assert or refrain from asserting p.

In fact we have already argued for this consequence of the strengthened determinacy role in 7.2.3. Suppose that it is determinate that I am in a context in which I have a reason to assert whether or not p, and that out of the relevant facts I am as knowledgeable as I can be. If indeterminacy is the only barrier to knowledge (i.e. if second order indeterminacy (and higher) is not a barrier as well) then this amounts to saying that I am only ignorant in the indeterminate cases.[23] In accordance with section 7.2.3, we have that, determinately:

> It's permissible to assert that p if and only if it is determinate that p

[23]The strengthened Δ-role stipulates that if it's determinate that p I could have known that p.

By fairly uncontroversial reasoning – namely, that Δ obeys the K axiom from §7.3.1 – it follows that if it is indeterminate whether it's determinate that p it is indeterminate whether you may assert that p.

What should I do in this situation? Should I assert or shouldn't I? I am suggesting there simply is no answer to these questions, there is no fact of the matter, and it would be misguided to continue asking. One could quite rightly imagine getting exasperated at this advice; after all, you have to do *something*, assert or do nothing, in that situation and knowing that there's no matter of fact which option I should pursue is not the least bit helpful.

Although this is clearly a frustrating situation to find oneself in I think that it is no more esoteric than more humdrum cases of vagueness in normative claims. Many normative properties are vague. One might for example, think that it is wrong to kill a person but not an embryo that is not a person. No doubt it is vague at which point one begins to be a person, and thus, according to our toy ethical theory, it will be vague at which point it ceases to be permissible to abort a child. Similar points apply to the norms of assertion. Consider a Sorites sequence for the property of being bald. I take it that it is permissible to assert that people with no hairs at all are bald, and that it is not permissible to assert people with 1000000 hairs are bald, but surely there will be cases, people with a certain distribution and number of hair, where it is borderline whether it's permissible to assert that they're bald. If Harry is such a case then there's simply no fact of the matter whether it's permissible or impermissible to assert that Harry is bald.

Another potential source of discomfort about the current theory might stem from the fact that the conditions for proper assertion are not always known to the subject. Surely, the objection might go, the conditions for proper assertion should be transparent, in the sense that both conditions for proper assertion, and the conditions for improper assertion, are luminous:

> If it is permissible for A to assert that p, then A is in a position to know that it is permissible for A to assert that p
>
> If it is not permissible for A to assert that p, then A is in a position to know that it is not permissible for A to assert that p

Surely, the objection might continue, it must be possible to operationalise assertion: one should be able to provide rules which tell you when it is and isn't ok to assert such that one is in principle always in a position to know whether one is complying with these rules.

I do not find this objection convincing. What the Sorites sequence above shows is that, quite independently, there will cases where there it is indeterminate whether you may assert p or not. Since indeterminacy bars knowledge these will be cases where you are not in a position to know whether p is assertible or not. In [144] Williamson criticises the view that normative rules should be operationalisable in the context of epistemology. I think much of what Williamson says about epistemology when your evidence is 'inexact', cases where you plausibly fail to know what your evidence is because it is vague what your evidence is, must apply to the epistemology of the paradoxes too.

7.3.5 Assertive uttering

We have focused our attention on the verb 'assert' which takes a that-clause in its right argument and relates a person to a proposition. There is also an important relation that holds between a person and a sentence when that person makes a certain kind of noise, perhaps with certain kinds of intentions. I shall call this 'assertive uttering.' These two relations are clearly very different. One asserts *by* assertively uttering a sentence, but the relationship is complicated by the fact that different sentences can be used to assert the same thing, and the same sentence in different languages and different contexts can be used to assert different things. Since one might worry that there are paradoxes that can

be formulated using the notion of assertive utterance which cannot be formulated using assertion it is worth addressing this issue.

The most important rule governing assertion that we have discussed is the norm

R. Do not assert what is unclear/indeterminate.

Like the rule 'do not assert p if you do not know that p' it provides an objective standard for evaluating assertions, one that is not always available to the asserter. In general there is no simple rule which would guarantee that you conformed to R, and would be such that you could always know that you're following the rule. We do not always know whether or not we know p, and we do always not know whether or not p is unclear. The fact that R (or the knowledge norm) provides an external standard of assessment means that sometimes it is impossible for *anyone* to know whether someone is complying with R (or the knowledge norm.) When it's unclear whether you know p or unclear whether it's clear that p (due to vagueness, perhaps, or the kind of second order unclarity the paradoxes generate) then it is impossible for anyone to know if someone has asserted p in conformity with R, or with the knowledge norm.

An interesting observation about the rules of assertion is that they are often language independent. For example the knowledge norm, R, the maxim 'be relevant', and other Gricean maxims appear to be generally applicable whatever your mode of communication is: they apply equally, whether you are speaking German, French or what have you.

On the other hand, when deciding which sentence to assertively utter in a given context, (and a given language, L) you must somehow incorporate your knowledge about what it would be appropriate to *assert* in that context, as determined by Gricean norms of conversation, R, and so on, with your particular linguistic knowledge telling you which sentence of L would result in you achieving this assertion. In order to have a theory of proper and improper assertive uttering, then, we must appeal to the speaker's *linguistic* knowledge in a way that we did not need to for our theory of proper assertion.

Here it is natural to think that we need specific linguistic knowledge about which sentences mean which propositions; if you have determined, via the very general norms of assertion that p would be the best proposition to assert we then we should strive to find a sentence of L which expresses p and assertively utter it. Even then there will be multiple sentences within that language that express p, some sentences express different things in different contexts, so there is plenty more to be said about this. The rules for appropriate assertive utterances are therefore much more complicated and less susceptible to systematic theorising than the rules for assertion. However, it is natural to think that something like this story is correct. In which case we can partly reduce the theory of assertive uttering to our theory of assertion, as governed by the principle R amongst other things. Thus:

> Assertively utter s in L at context c only if you know that s says that p in L at c and it's permissible to assert p.

It is natural, therefore, to think that there would be no special problem in constructing a theory which contains its own theory of assertive uttering, as well as assertion. The expressive requirements needed seem to be (i) that your theory can state when it is and isn't permissible to assert p (see section 7.2.3) and (ii) that the theory contains a 'saying' connecticate, which states when a sentence means that p. Both of these are easy to accommodate.[24]

The important thing to note about this reduction is that in assertively uttering 'p' in English you have not always asserted that p. This connection relies on the contingent fact that 'p' means that p in English. But the principle can fail when 'p' contains indexical expressions, and most importantly, when 'p' contains semantic vocabulary. You would quickly find yourself confronted with paradoxes if you simply translated principles involving

[24]We can prove the consistency of a very basic theory of saying by interpreting the connecticate 's says that p' in DFS, by defining it as 's is true iff p'.

the locution 'it's permissible to assert that p' with the locution 'it's permissible to assertively utter 'p"; these only approximate one another when 'p' says that p in English.

7.4 A semantically self-contained language

7.4.1 The consistency of DFS

The consistency of DFS is shown by a model theoretic construction.

A Kripke model for the language $\mathcal{L}$ is a quadruple $\langle W, R, D, [\![\cdot]\!].\rangle$ consisting of a set of worlds, W, an accessibility relation, R, a domain, D, and interpretation function $[\![\cdot]\!]_w$ which assigns each term, function and predicate symbol an element, function, or relation over D of the respective type, relative to each world w.

In our model both W and D will be the set of natural numbers, Rxy the relation that holds if $x = y$ or $x = y + 1$, and each primitive piece of arithmetical vocabulary is assigned its standard interpretation relative to every world, for example $[\![+]\!]_w = \{\langle\langle x, y\rangle, x + y\rangle \mid x, y \in \mathbb{N}\}$, $[\![0]\!]_w = 0$, etc, for every $w \in W$. I shall denote the element $n \in W$ as w_n to distinguish it from the same element n in the domain of natural numbers D. An assignment function, v, is a function from variables of $\mathcal{L}$ to D.

All that is left is to specify the extension of Tr at each world w_i, $[\![Tr]\!]_{w_i}$. We shall do this in a moment.

We can define what it means for a formula of $\mathcal{L}$ relative to an assignment to hold at a world in a Kripke model of this kind, which we can write $w_n, v \Vdash \phi$, or just $w_n \Vdash \phi$ if ϕ is closed. I shall omit the details of the definition (see the next section.) What is important is that whether a formula holds at a world, w, depends only on whether the sentence holds at worlds which are R accessible to w, or R accessible to worlds which are R accessible to w, and so on. In our model this just amounts to this: whether ϕ holds at w_n depends only on whether or not ϕ holds at worlds w_i for $i \leq n$. Moreover, whether ϕ holds at w_n never depends on whether it holds at w_{n+1}. Thus following definition of the extension of the truth predicate at a world w_n is not viciously circular:

$$[\![Tr]\!]_{w_0} = \emptyset$$

$$[\![Tr]\!]_{w_{n+1}} := \{\ulcorner\phi\urcorner \mid \phi \text{ is closed and } w_n \Vdash \phi\}$$

To get a picture of the construction see the diagram below

$$\emptyset \qquad \{\ulcorner\phi\urcorner \mid w_0 \Vdash \phi\} \qquad \{\ulcorner\phi\urcorner \mid w_1 \Vdash \phi\} \qquad \{\ulcorner\phi\urcorner \mid w_2 \Vdash \phi\} \ldots$$

All the axioms of DFS hold at w_1. In fact every theorem of DFS_n holds at w_{n+1}. It follows that DFS_n has a standard model, w_{n+1}, for each n.

Say that a sentence eventually holds in this model if there exists an i such that the sentence holds at every world w_j for $j > i$. The consistency of DFS is established by noting that every theorem holds at every world to the right of w_1, and therefore eventually holds, and that if ϕ eventually holds, so does $Tr(\ulcorner\phi\urcorner)$ and $\Delta\phi$ and if $Tr(\ulcorner\phi\urcorner)$ eventually eventually holds so does ϕ. Furthermore the classical rules of inference preserve eventual holding, so every theorem of DFS eventually holds. The consistency of DFS is obtained by noting that $1 = 0$ does not eventually hold.

7.5 Semantic self-sufficiency

It cannot be stressed enough that the model we provided in the previous section is not the intended intepretation of $\mathcal{L}$. Its purpose there is to establish syntactic facts about certain theories – that no contradiction is provable in DFS, and that no contradiction is provable in DFS_n given the ω-rule. Since no consistent recursively enumerable theory can prove its own consistency this model is, speaking loosely, out of reach from inside the language. Either $\mathcal{L}$ is not rich enough to describe this model, or, if it is, the theory DFS is not strong enough to prove that the axioms hold in this model and that its rules are preserved.

Nonetheless the intended interpretation of the language presumably can be taken to have the form of a Kripke model, consisting of a set or class $\mathcal{I}$ of indices and an accessibility relation, in which the truth conditions for a statement of the form 'determinately p' are given by the truth of 'p' at all indices accessible from the actual index.

In our understanding indices are maximally consistent propositions. The notion of consistency at play is not necessarily metaphysical possibility, therefore the indices may or may not represent possible worlds. If i is an index, i.e. a maximally consistent proposition, we can say that the proposition p holds at i, written $i \models p$, just in case i entails p. According to this understanding it strictly makes no sense to talk about a sentence being true at an index. We therefore analyse the relation 'ϕ' is true at i as simply meaning that the proposition that 'ϕ' is true holds at i, in formalism: $i \models Tr(\ulcorner\phi\urcorner)$.[25] We shall introduce the shorthand $i \Vdash \ulcorner\phi\urcorner$ for $i \models Tr(\ulcorner\phi\urcorner)$.

In the more general setting we will want to talk of a sentence being true at an index relative to an assignment. We can eliminate this need in the arithematical setting – we can give the truth of a quantified statement in terms of the truth of all of its substitution instances eliminating the need for a satisfaction relation (since they are interdefinable in the arithmetical setting nothing of importance rests on this.) Assume that i ranges over indices in $\mathcal{I}$. We need a theory of truth at an index

- '$\phi \wedge \psi$' is true at i if and only if 'ϕ' is true at i and 'ψ' is true at i.
- '$\phi \vee \psi$' is true at i if and only if 'ϕ' is true at i or 'ψ' is true at i.
- '$\neg\phi$' is true at i if and only if 'ϕ' is not true at i.
- '$\forall x\phi$' is true at i if and only if '$\phi[t/x]$' is true at i for every numeral term 't'.
- '$\exists x\phi$' is true at i if and only if '$\phi[t/x]$' is true at i for some numeral term 't'.
- '$\Delta\phi$' is true at i if and only if for every j accessible from i, 'ϕ' is true at j.

This is a standard Kripke style semantics which determines how the truth of a sentence at an index depends on the truth of its parts. For arithmetical atomic sentences we can say:

- For all indices, i, 'ϕ' is true at i if 'ϕ' is a true arithmetical sentence.

The preceding principles encapsulate a number of things that I know about the intended interpretation. There are also a number of things that I do not know about it. For example, I do not know whether the Gödel number of the statement of Goldbach's conjecture is in the extension of Tr or not. The example of Goldbach's conjecture demonstrates that a semantic theory should not be attempting to aim at a complete description of the intended model. It is impossible to knowledgeably state a theory which complete describes the intended model because we simply do not know whether what the extension of all the predicates are. What makes it particularly difficult is when the language contains vague or indeterminate

[25] We must think of a sentence being true as meaning roughly, that the sentence is true as we in fact use it, which need not necessarily be truth according to how it is used at the index in question. Really, what we want to say is that a sentence 'ϕ' is true at an index if the proposition 'ϕ' (actually) expresses is entailed by i, but we cannot do that unless we have the expressing relation in the language (see chapter 6.)

vocabulary: I do not know whether or not the liar sentence is true in the intended model because it is indeterminate whether it is true in the intended model. Thus one might have noted that there is no clause for atomic sentences involving the truth predicate in this theory. The naïve theory of truth might say

- '$Tr(\ulcorner\phi\urcorner)$' is true at i if and only if 'ϕ' is true at i.

This would have the effect of giving the truth predicate a disquotational meaning relative to every index and would lead to paradoxes. We can state partial truths, such as the fact '1=1' is in the truth predicate, as is every true atomic arithmetical sentence. Via the compositionality clauses this determines the truth or falsity of every arithmetical sentence. However we cannot in general say whether the truth-teller or the liar are true.

So much for truth at an index. Finally we can say that a sentence is true just in case it is true at the actual index, i^*.

Our next goal is to extend the language of DFS so that it can state this semantics in such a way that it applies to itself. (For further discussion of why we might want to do this see Fitch's discussion of 'universal metalanguages' [45].) In the purely extensional theory FS, obtained by removing the determinacy axioms from DFS (see Halbach [55]), this project is possible since the theory contains its own truth predicate, and, since the theory is arithmetical, the satisfaction relation is reducible to truth. However in DFS we have the non-extensional connective Δ to deal with so the semantics is not extensional.

In order to formalise such a theory we must add some new vocabulary to the language of DFS. The theory we develop identifies indices with natural numbers; nothing turns on this, but it makes the theory a lot simpler to state.

> A predicate, Ix, saying that x is an index.
>
> A relation $i \Vdash x$ relating an index (represented by a natural number) on the left, to the Gödel number of a sentence.
>
> A two place relation Rxy, for the accessibility relation.
>
> A constant, i^*, denoting the actual index.

(Aside: it would be easy to also add the $\models$ relation to the language, this time represented as a connecticate taking a term, an index/natural number on the right, with a term.)

Call this language $\mathcal{L}^+$. Call the result of adding the following axioms to DFS DFS$^+$:

- $Ii \rightarrow (i \Vdash \ulcorner\phi \wedge \psi\urcorner \leftrightarrow (i \Vdash \ulcorner\phi\urcorner \wedge i \Vdash \ulcorner\psi\urcorner))$
- $Ii \rightarrow (i \Vdash \ulcorner\phi \vee \psi\urcorner \leftrightarrow (i \Vdash \ulcorner\phi\urcorner \vee i \Vdash \ulcorner\psi\urcorner))$
- $Ii \rightarrow (i \Vdash \ulcorner\neg\phi\urcorner \leftrightarrow \neg i \Vdash \ulcorner\phi\urcorner)$
- $Ii \rightarrow (i \Vdash \ulcorner\forall x\phi\urcorner \leftrightarrow \forall y i \Vdash \ulcorner\phi[\dot{y}/x]\urcorner)$
- $Ii \rightarrow (i \Vdash \ulcorner\exists x\phi\urcorner \leftrightarrow \exists y i \Vdash \ulcorner\phi[\dot{y}/x]\urcorner)$
- $Ii \rightarrow (i \Vdash \ulcorner\Delta\phi\urcorner \leftrightarrow \forall j(Rij \rightarrow j \Vdash \ulcorner\phi\urcorner))$
- $Ii \wedge At_{PA}(\ulcorner\phi\urcorner) \rightarrow (i \Vdash \ulcorner\phi\urcorner \leftrightarrow ver(\ulcorner\phi\urcorner))$

These axioms are simply formalisations of the definition of truth at a world we specified earlier.

Note that we also have new vocabulary which we need to give semantics for. We can do this as follows:

- $i \Vdash \ulcorner j \Vdash \phi\urcorner \leftrightarrow j \Vdash \phi$
- $i \Vdash \ulcorner Rst\urcorner \leftrightarrow Rst$

One might also want an axiom stating that what the indices are is index invariant, i.e.: 'j is an index' is true at i iff j is and index. The model we construct does not validate this principle. Which numbers are indices is an indeterminate matter since we can formulate paradoxes involving the notion of being true at all indices.[26] Note also that it's indeterminate which index i^* refers to.

The consistency of DFS^+ is shown using the same model construction we mentioned above. Since we have overloaded our use of the symbol $\Vdash$ I shall use $\Vdash^*$ to denote the metatheoretic notion we're using in the consistency proof, and save $\Vdash$ for the object language relation.

$\emptyset$ $\quad$ $\{\phi \mid w_0 \Vdash^* \phi\}$ $\quad$ $\{\phi \mid w_1 \Vdash^* \phi\}$ $\quad$ $\{\phi \mid w_2 \Vdash^* \phi\} \ldots$

We augment the model above as follows:

$[\![I]\!]_{w_i} := \{j \mid \lfloor \frac{i}{2} \rfloor \leq j \leq i\}$

$[\![R]\!]_{w_i} := \{\langle j, k\rangle \mid j = k + 1 \text{ or } j = k\}$

$[\![\Vdash]\!]_{w_i} := \{\langle k, n\rangle \mid \lfloor \frac{i}{2} \rfloor \leq k \leq i, n \in [\![Tr]\!]_{w_k}\}$

$[\![i^*]\!]_{w_i} := i$

[26] Otherwise one could formulate paradoxes, such as I'm not true at all indices.

Bibliography

[1] A.J. Ayer. Language, truth and logic. 1936.

[2] J. Azzouni. The inconsistency of natural languages: how we live with it. *Inquiry*, 50(6):590–605, 2007.

[3] E. Barnes and J.R.G. Williams. A theory of metaphysical indeterminacy. *Oxford studies in metaphysics*, 6, 2010.

[4] D. Barnett. Is vagueness sui generis? 1. *Australasian Journal of Philosophy*, 87(1):5–34, 2009.

[5] D. Barnett. Indeterminacy and incomplete definitions. *Journal of Philosophy*, 105(4):167, 2010.

[6] D. Barnett. Does vagueness exclude knowledge? *Philosophy and Phenomenological Research*, 82(1):22–45, 2011.

[7] J. Barwise and J. Etchemendy. *The liar: An essay on truth and circularity*. Oxford University Press, USA, 1989.

[8] J. Beall. *Spandrels of truth*. Oxford University Press, USA, 2009.

[9] R.A. Benton. A Simple Incomplete Extension of T which is the Union of Two Complete Modal Logics with F.M.P. *Journal of Philosophical Logic*, 31(6):527–541, 2002.

[10] P. Blackburn, M. De Rijke, and Y. Venema. *Modal logic*. Cambridge University Press, 2002.

[11] G. Boolos. To be is to be a value of a variable (or to be some values of some variables). *The Journal of Philosophy*, 81(8):430–449, 1984.

[12] G. Boolos. Nominalist platonism. *The Philosophical Review*, 94(3):327–344, 1985.

[13] R. Bradley. Conditional desirability. *Theory and decision*, 47(1):23–55, 1999.

[14] R. Brandom. The significance of complex numbers for frege's philosophy of mathematics. In *Proceedings of the Aristotelian Society*, volume 96, pages 293–315, 1996.

[15] D. Braun and T. Sider. Vague, so untrue. *Noûs*, 41(2):133–156, 2007.

[16] T. Burge. Individualism and the mental. *Midwest studies in philosophy*, 4(1):73–121, 1979.

[17] J.A. Burgess. The sorites paradox and higher-order vagueness. *Synthese*, 85(3):417–474, 1990.

[18] R. Carnap. *The logical syntax of language*. London: Routledge & Kegan Paul, 1937.

[19] B.F. Chellas. *Modal logic: an introduction*. Cambridge Univ Pr, 1980.

[20] C. Chihara. The semantic paradoxes: A diagnostic investigation. *The Philosophical Review*, 88(4):590–618, 1979.

[21] D. Deutsch. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London*, 1999.

[22] Harry Deutsch. Diagonalization and truth functional operators. *Analysis*, 70(2):215–217, 2010.

[23] R. Dietz. Betting on borderline cases. *Philosophical Perspectives*, 22(1):47–88, 2008.

[24] C. Dorr. Vagueness without ignorance. *Philosophical Perspectives*, 17(1):83–113, 2003.

[25] C. Dorr. How Vagueness Could Cut Out At Any Order. *Unpublished*, 2010.

[26] C. Dorr. Iterating Definiteness. *Cuts and Clouds*, 1(9):550–586, 2010.

[27] D. Edgington. Vagueness by degrees. *Vagueness: A reader*, pages 294–316, 1996.

[28] M. Eklund. Inconsistent languages. *Philosophy and Phenomenological Research*, 64(2):251–275, 2002.

[29] M. Eklund. What vagueness consists in. *Philosophical Studies*, 125(1):27–60, 2005.

[30] A. Elga. Subjective probabilities should be sharp. *Philosophers Imprint*, 10, 2010.

[31] J. Fantl and M. McGrath. *Knowledge in an uncertain world*. Oxford University Press Oxford, 2009.

[32] D.G. Fara. An anti-epistemicist consequence of margin for error semantics for knowledge. *Philosophy and Phenomenological Research*, 64(1):127–142, 2002.

[33] D.G. Fara. Gap principles, penumbral consequence, and infinitely higher-order vagueness. *Liars and Heaps: New Essays on Paradox*, pages 195–222, 2003.

[34] S. Feferman. Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56(1):1–49, 1991.

[35] H. Field. A note on jeffrey conditionalization. *Philosophy of Science*, pages 361–367, 1978.

[36] H. Field. Some thoughts on radical indeterminacy. *The Monist*, 81(2):253–273, 1998.

[37] H. Field. Indeterminacy, degree of belief, and excluded middle. *Nous*, 34(1):1–30, 2000.

[38] H. Field. Attributions of meaning and content. *Truth and the Absence of Fact*, 1(9):157–175, 2001.

[39] H. Field. The semantic paradoxes and the paradoxes of vagueness. *Liars and Heaps*, pages 262–311, 2003.

[40] H. Field. *Saving truth from paradox*. Oxford University Press, USA, 2008.

[41] H. Field. This magic moment: Horwich on the boundaries of vague terms. *Cuts and Clouds: Vagueness, its Nature and its Logic. Oxford University Press, Oxford*, 2010.

[42] K. Fine. Propositional quantifiers in modal logic1. *Theoria*, 36(3):336–346, 1970.

[43] K. Fine. Vagueness, truth and logic. *Synthese*, 30(3):265–300, 1975.

[44] K. Fine. The Impossibility of Vagueness. *Philosophical Perspectives*, 22(1):111–136, 2008.

[45] F.B. Fitch. Universal metalanguages for philosophy. *The Review of Metaphysics*, pages 396–402, 1964.

[46] G. Frege. Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. *Reprinted in [9]*, 1979.

[47] D.M. Gabbay and V.B. Shehtman. Products of modal logics, part 1. *Logic journal of IGPL*, 6(1):73–146, 1998.

[48] H. Gaifman. Pointers to truth. *The Journal of Philosophy*, 89(5):223–261, 1992.

[49] D. Garber. Field and jeffrey conditionalization. *Philosophy of Science*, 47(1):142–145, 1980.

[50] D. Graff. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, 28(1):45–82, 2002.

[51] H. Greaves. Understanding deutsch's probability in a deterministic multiverse. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 35(3):423–456, 2004.

[52] H.P. Grice. Meaning. *The philosophical review*, 66(3):377–388, 1957.

[53] D.L. Grover. Propositional quantifiers. *Journal of Philosophical Logic*, 1(2):111–136, 1972.

[54] A. Gupta and N.D. Belnap. *The revision theory of truth*. The MIT Press, 1993.

[55] V. Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, 35(3):311–327, 1994.

[56] J. Hawthorne. Epistemicism and semantic plasticity. *Metaphysical essays*, 1(9):185–211, 2006.

[57] H.G. Herzberger. Truth and modality in semantically closed languages. *The Paradox of the Liar, Yale University Press, New Haven*, pages 25–46, 1970.

[58] H.G. Herzberger. New paradoxes for old. In *Proceedings of the Aristotelian Society*, volume 81, pages 109–123. JSTOR, 1980.

[59] H.G. Herzberger. Naive semantics and the liar paradox. *The Journal of Philosophy*, 79(9):479–497, 1982.

[60] P. Horwich. *Truth*, volume 8. Dartmouth Publishing company, 1994.

[61] P. Horwich. The nature of vagueness. *Philosophy and Phenomenological Research*, 57(4):929–935, 1997.

[62] R. Jeffrey. The sure thing principle. In *Proceedings of the biennial meeting of the philosophy of science association*, pages 719–730, 1982.

[63] R.C. Jeffrey. *The logic of decision*. University of Chicago Press, 1990.

[64] J.M. Joyce. A nonpragmatic vindication of probabilism. *Philosophy of Science*, pages 575–603, 1998.

[65] J.M. Joyce. *The foundations of causal decision theory*. Cambridge Univ Pr, 1999.

[66] H. Kamp. Two theories about adjectives. *Formal semantics of natural language*, pages 123–155, 1975.

[67] D. Kaplan. A problem in possible-world semantics. *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, pages 41–52, 1995.

[68] D. Kaplan and R. Montague. A paradox regained. *Notre Dame journal of formal logic*, 1(3):79–90, 1960.

[69] S. Kearns and O. Magidor. Semantic sovereignty. *forthcoming in Philosophy and Phenomenological Research*.

[70] S. Kearns and O. Magidor. Epistemicism about vagueness and meta-linguistic safety. *Philosophical Perspectives*, 22(1):277–304, 2008.

[71] R. Keefe. *Theories of vagueness*. Cambridge University Press, 2000.

[72] M. Kölbel. Vagueness as semantic. *Cuts and clouds: vagueness, its nature, and its logic*, page 304, 2010.

[73] S. Kripke. Outline of a theory of truth. *The journal of philosophy*, 72(19):690–716, 1975.

[74] S. Kripke. Speaker's reference and semantic reference. *Midwest studies in philosophy*, 2(1):255–276, 1977.

[75] S.A. Kripke. A puzzle about time and thought.

[76] G. Küng. The meaning of the quantifiers in the logic of leśniewski. *Studia logica*, 36(4):309–322, 1977.

[77] A. Kurucz. 15 combining modal logics. *Studies in Logic and Practical Reasoning*, 3:869–924, 2007.

[78] M. Lasonen-Aarnio. Unreasonable knowledge. *Philosophical Perspectives*, 24(1):1–21, 2010.

[79] H. Leitgeb. What truth depends on. *Journal of philosophical logic*, 34(2):155–192, 2005.

[80] D. Lewis. General semantics. *Synthese*, 22(1):18–67, 1970.

[81] D. Lewis. A subjectivists guide to objective chance. *Studies in inductive logic and probability*, 2:263–293, 1980.

[82] D. Lewis. Logic for equivocators. *Noûs*, 16(3):431–441, 1982.

[83] D. Lewis. Meaning without use: Reply to hawthorne. 1992.

[84] D.K. Lewis. *Convention: A philosophical study*. Wiley-Blackwell, 2002.

[85] J. MacFarlane. The things we (sorta kinda) believe. *Philosophy and Phenomenological Research*, 73(1):218–224, 2006.

[86] J. MacFarlane. Fuzzy epistemicism. *Cuts and Clouds: Vagueness, its Nature and its Logic*, pages 438–463, 2010.

[87] A. Mahtani. Can Vagueness Cut Out at Any Order? *Australasian Journal of Philosophy*, 86(3):499–508, 2008.

[88] N. Malcolm. Are necessary propositions really verbal? *Mind*, 49(194):189–203, 1940.

[89] T. Maudlin. *Truth and paradox: solving the riddles*. Oxford University Press, 2004.

[90] T. Maudlin. Reducing revenge to discomfort. *Revenge of the Liar*, page 184, 2007.

[91] V. McGee. *Truth, vagueness, and paradox: An essay on the logic of truth*. Hackett Publishing Company Inc, 1990.

[92] V. McGee and B. McLaughlin. Distinctions without a difference. *The Southern Journal of Philosophy*, 33(S1):203–251, 1995.

[93] D. Mirimanoff. Les antinomies de russell et de burali-forti: et le problème fondamental de la théorie des ensembles, 1917.

[94] J.D. Monk and R. Bonnet. *Handbook of Boolean algebras*, volume 2. North Holland, 1989.

[95] R. Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta philosophica fennica*, 16:153–167, 1963.

[96] J. Myhill. A refutation of an unjustified attack on the axiom of reducibility. *Bertrand Russell Memorial Volume*, pages 81–90, 1979.

[97] R. Nozick. Newcombs problem and two principles of choice. *Rationality in action: Contemporary approaches*, pages 207–234, 1990.

[98] D. Parfit. Personal identity. *The Philosophical Review*, 80(1):3–27, 1971.

[99] C. Parsons. The liar paradox. *Journal of Philosophical Logic*, 3(4):381–412, 1974.

[100] D. Patterson. Inconsistency theories of semantic paradox. *Philosophy and Phenomenological Research*, 79(2):387–422, 2009.

[101] G. Priest. Review of truth and paradox: Solving the riddles. *Journal of philosophy*, 102(9):483–486, 2005.

[102] G. Priest. *In contradiction: a study of the transconsistent*. Oxford University Press, USA, 2006.

[103] G. Priest. Revenge, field, and zf. *Revenge of the Liar*, page 225, 2007.

[104] G. Priest. Hopes fade for saving truth. *Philosophy*, 85(1):109, 2010.

[105] GRAHAM PRIEST. Review of truth, vagueness and paradox. *Mind*, 103(411):387–391, 1994.

[106] A. Prior. On a family of paradoxes. *Notre Dame Journal of Formal Logic*, 2(1):16–32, 1961.

[107] A.N. Prior. Definitions, rules and axioms. In *Proceedings of the Aristotelian Society*, volume 56, pages 199–216, 1955.

[108] A.N. Prior. *Changes in events and changes in things*. 1962.

[109] A.N. Prior and AN Prior. *Objects of thought*. Oxford, 1971.

[110] D. Raffman. Vagueness without paradox. *The Philosophical Review*, 103(1):41–74, 1994.

[111] D. Raffman. Demoting higher-order vagueness. *Cuts and Clouds*, pages 509–522, 2010.

[112] F.P. Ramsey and G.E. Moore. Facts and propositions. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 7:153–206, 1927.

[113] A. Rayo. A metasemantic account of vagueness. *Cuts and Clouds: Vagueness, Its Nature, and Its Logic, eds R. Dietz and S. Moruzzi*, pages 23–45, 2010.

[114] A. Rayo and S. Yablo. Nominalism through de-nominalization. *Noûs*, 35(1):74–92, 2001.

[115] B. Russell. *The principles of mathematics*. WW Norton & Company, 1903.

[116] R.M. Sainsbury. Concepts without boundaries. *Vagueness: a reader*, pages 251–264, 1996.

[117] K. Scharp. Replacing truth. *Inquiry*, 50(6):606–621, 2007.

[118] S. Schiffer. Vagueness and partial belief. *Philosophical Issues*, 10(1):220–257, 2000.

[119] A. Sennet. Semantic plasticity and epistemicism. *Philosophical Studies*, pages 1–13, forthcoming.

[120] S. Shapiro. *Vagueness in context*. Oxford University Press, USA, 2006.

[121] T. Sider. Reductive theories of modality. *The Oxford handbook of metaphysics*, 2001, 2003.

[122] K. Simmons. *Universality and the liar: An essay on truth and the diagonal argument*. Cambridge Univ Pr, 1993.

[123] N.J.J. Smith. Degree of belief is expected truth value. *Cuts and Clouds*, 1(9):491–507, 2010.

[124] N.J.J. Smith. Vagueness, uncertainty and degrees of belief: Two kinds of indeterminacy; one kind of credence. *Erkenntnis*, forthcoming.

[125] S. Soames. *Understanding truth*. Oxford University Press, USA, 1999.

[126] R. Stalnaker. A defense of conditional excluded middle. *Ifs: Conditionals, belief, decision, chance, and time*, pages 87–104, 1981.

[127] R. Stalnaker. Responses to stanley and schlenker. *Philosophical studies*, 151(1):143–157, 2010.

[128] R.C. Stalnaker. *Inquiry*. 1984.

[129] J. Stanley. Assertion and intentionality. *Philosophical studies*, 151(1):87–113, 2010.

[130] A. Tarski. The concept of truth in formalized languages. *Logic, semantics, metamathematics*, pages 152–278, 1956.

[131] A. Tarski. Truth and proof. *Scientific American*, 220:66, 1969.

[132] P. Unger. There are no ordinary things. *Synthese*, 41(2):117–154, 1979.

[133] D. Wallace. Everettian rationality: defending deutsch's approach to probability in the everett interpretation. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):415–439, 2003.

[134] J.R.G. Williams. Classicism and indeterminacy.

[135] J.R.G. Williams. Indeterminate survival: by degrees.

[136] J.R.G. Williams. Degree supervaluational logic. *The Review of Symbolic Logic*, 4(01):130–149, 2011.

[137] T. Williamson. Inexact knowledge. *Mind*, 101(402):217–242, 1992.

[138] T. Williamson. *Vagueness*. Routledge, 1994.

[139] T. Williamson. Definiteness and knowability. *The Southern journal of philosophy*, 33(S1):171–192, 1995.

[140] T. Williamson. Imagination, stipulation and vagueness. *Philosophical Issues*, 8:215–228, 1997.

[141] T. Williamson. Indefinite extensibility. *Grazer Philosophische Studien*, pages 1–24, 1998.

[142] T. Williamson. On the structure of higher-order vagueness. *Mind*, 108(429):127, 1999.

[143] T. Williamson. *Knowledge and its Limits*. Oxford University PRess, 2000.

[144] T. Williamson. Why epistemology cant be operationalized. *Epistemology: New Philosophical Essays*, 2008.

[145] L. Wittgenstein. *Tractatus logico-philosophicus*. 1922.

[146] C. Wright. Is higher order vagueness coherent? *Analysis*, 52(3):129, 1992.

[147] C. Wright. On being in a quandary. relativism vagueness logical revisionism. *Mind*, 110(437):45–98, 2001.

[148] C. Wright. The illusion of higher-order vagueness. *Cuts and Clouds*, pages 523–549, 2010.

[149] S. Yablo. Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11(1):117–137, 1982.

[150] S. Yablo. Definitions, consistent and inconsistent. *Philosophical Studies*, 72(2):147–175, 1993.

[151] A.M. Yaqūb. *The liar speaks the truth: A defense of the revision theory of truth*. Oxford University Press, USA, 1993.

[152] E. Zardini. Squeezing and stretching: How vagueness can outrun borderlineness. In *Proceedings of the Aristotelian Society (Hardback)*, volume 106, pages 421–428. John Wiley & Sons, 2006.

[153] M.J. Zimmerman. Propositional quantification and the prosentential theory of truth. *Philosophical Studies*, 34(3):253–268, 1978.