Shifty Semantics and Slippery Slopes

Essays on semantic plasticity and the epistemic theory of vagueness

Aleksander Domosławski

Submitted in accordance with the requirements for the degree of Doctor of Philosophy.

The University of Leeds
School of Philosophy, Religion and History of Science

August 2022

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

To my parents, Małgorzata and Artur

To my wife, Natalia

Acknowledgements

This work would not be possible without the help and support of family, friends and teachers I encountered throughout the years. I am truly grateful.

First of all, I would like to thank my wonderful supervisors Daniel Elstein and Robbie Williams. I could not ask for better advisors. I am grateful they read through my half-baked ideas and were always supportive even though at times their well-placed objections caught me like a deer in the headlights. Both of their contributions have made the project significantly better. Daniel's encouragement and willingness to discuss my work (some meetings lasting up to six hours) gave me the invaluable sense that I have something interesting to say. Weekly (or biweekly) meetings with Robbie have been indispensable to keeping me on track with the work during the pandemic. Robbie's unparalleled ability to parse out my writing and ramblings during meetings and present me with the most interesting version of my own work has been greatly encouraging. I am more open-minded and courageous in my philosophical thinking due to reading Robbie's work and discussing my work with him.

I am grateful to the funders of my project. This work was supported by the Arts & Humanities Research Council (grant number AH/L503848/1), the School of Philosophy, Religion and History of Science at the University of Leeds and the Royal Institute of Philosophy.

I have had the good fortune to encounter many great teachers throughout the years. Many thanks to Quassim Cassam, whose lectures on the history of modern philosophy have got me permanently hooked on philosophy, even though I focus on Kant significantly less than I thought at the time. Thanks to Filip Kawczyński and Paul Elbourne who pushed me away from Kant towards the topics that occupy me today. Last but not least, I am greatly indebted to Tim Williamson who supervised my master's thesis, where some of the ideas developed in this project have first begun to take shape. Tim's influence on my thinking in general and this thesis in particular, both through my conversations with him and reading his work, should be fairly obvious.

I am grateful to my philosopher friends. Thanks to the Botany House crew: Adina Covaci, Alec Siantonas, Alexios Stamatiadis-Bréhier, Andreas Bruns, Arthur Carlyle, David Heering, Miriam Bowen, Radu Bumbăcea, and Sam Mason.

Special thanks go to my brother Ignacy, Michał Maśny, Miriam Bowen and Tomek Zyglewicz who read significant portions of this thesis. My work has benefitted greatly from hearing what they have to say about it. I am also grateful to Michał and Tomek for the many philosophical conversations throughout the years and reminding me that philosophy does not begin and end with epistemicism.

My warmest thanks go to Ala Smołkowska, Łukasz Adamski, Michał Strzałkowski, Natalia Sarnacka, Olaf Sota, Tomek Fosiewicz, and Tosia Oraczewska, who helped me survive the isolation during the pandemic. I am also grateful to Filip Witek and Magda Płotczyk for providing safe havens during my trips to London. A special nod goes to my friends Andrzej Kurbiel, Marek Rymarz, and Mikołaj Olszewski who occasionally let me beat them at chess to take my mind off philosophy.

I am eternally grateful to my parents and my brother for their love and support. Thanks to Ignacy for the many laughs over the years. Thanks to my mother Małgorzata, who was always willing to make the effort to understand what I was working on and alleviate any existential angst. Thanks to my father Artur for the advice and empathy with the struggles of writing throughout the years. You guys are the best.

Finally, my deepest gratitude goes to my wife Natalia. In difficult moments she always reminds me of what's really important. Plus the constant excitement she brings to my life is the perfect counterweight to the slightly monotonous academic work.

Abstract

Our words have meaning. What meanings they possess is determined by meaning-fixing factors, e.g. usage. Large changes in the meaning-fixing facts produce large shifts in meaning. It's natural to conjecture that even the tiniest changes in usage produce tiny shifts in meaning - after all, any large change can be broken down into a series of small changes. This conjecture is known as *semantic plasticity*.

Semantic plasticity is the central component of the epistemicist theory of vagueness. Furthermore, several puzzling consequences of semantic plasticity have recently been identified. This thesis is the first sustained examination of semantic plasticity that articulates its motivation, refines its characterization, and answers this battery of recent puzzles. This work is essential if Williamsonian epistemicism is to be tenable. But everyone needs to square the arguments for semantic plasticity with the puzzles it engenders, and so the work has wide relevance.

The aim of the thesis is threefold. Firstly, it sums up and clears up the debate engendered by the publication of *Vagueness* in 1994. Secondly, it refines the epistemicist account in the areas that were shown to be lacking in the literature. Thirdly, it answers some recent puzzles generated by semantic plasticity.

The first chapter sums up the debate about epistemicism and semantic plasticity over the last 30 years. Later chapters are divided into three parts. Part I defends and refines epistemicist metasemantics, including the metasemantics of moral terms. Part II develops the technical details of the theory, providing an account of the definiteness operator, a model for ignorance due to semantic plasticity (and other factors) and an account of the restrictions on metalinguistic vocabulary like 'truth' imposed by epistemicism. Part III presents an account of metalinguistic comparatives using some elements of the epistemicist machinery, and answers a recent puzzle generated by semantic plasticity.

Contents

Acknowledgements v

Abstract vii

Contents viii

Introduction 1

1 Vagueness and the Sorites 2

2 How to read the thesis 11

Chapter 1 Epistemicism: 30 years after 14

1 Introduction 14

2 Summary of the Epistemic View of Vagueness 16

3 Logical Issues 21

4 Metasemantic Issues 31

5 Epistemic Issues 36

6 Semantic Issues 48

7 Puzzle cases 50

8 Conclusion 52

PART I – METASEMANTICS 53

Chapter 2 Metasemantic Problems for Epistemicism 54

1 Introduction 54

2 Metasemantic Commitments of Epistemicism 56

3 Objections to (1) - sharpness of vague predicates 62

4 Objections to (2) 68

5 Conclusion 81

Chapter 3 Shifty Morals 82

1 Introduction 83

2 Preliminaries 85

3 Problems for epistemicism 88

4 Towards an epistemicist solution 90

5 The reason-responsiveness metasemantics for moral terms 95

6 The epistemicist solution to moral vagueness 98

7 Potential problems 105

8 Conclusion 110

PART II – TECHNICAL ISSUES 111

Chapter 4 Factor-relative Safety 113

1 Introduction 113

2 The safety condition on knowledge 114

3 Factor-relative safety 123

4 The simple model 124

5 The complex model 132

6 What to do with factor-relative safety? 138

7 Conclusion 140

Chapter 5 Defining Definiteness 142

1 Introduction 142

2 Multidimensional models 143

3 Safety and definiteness 147

4 Desiderata for an account of definiteness 149

5 Epistemicist Definiteness 153

6 Alternatives and their issues 162

7 Conclusion 171

Chapter 6 Vagueness of Truth 172

1 Introduction 173

2 The T-Schema Puzzle 174

3 Clarifying the puzzle 177

4 Consequences of the puzzle 181

5 Conclusion 188

PART III – PUZZLES 189

Chapter 7 The Omniscient Speaker Puzzle 190

1 Introduction 190

2 Epistemicism 191

3 The Omniscient Speaker Puzzle 192

4 Choice points for epistemicism 194

5 Conclusion 206

Chapter 8 Metalinguistic Comparatives and Semantic Plasticity 208

1 Introduction 209

2 Desiderata 210

3 The Similarity View 211

4 Existing frameworks 223

5 Conclusion 230

Concluding remarks 231

Bibliography **234**

Introduction

The idea for this project first arose when I was attending the legendary MLE[1] reading group during my time at Oxford. We were reading a series of papers on epistemicism and semantic plasticity, most notably articles by Ofra Magidor and John Hawthorne, and I was captivated by the issues they presented. When reading these papers I had the feeling that solutions must be right around the corner and yet at the same time they were nowhere to be found. I didn't have a clearly outlined idea for a project then, but I knew that I wanted to work on these problems and with some luck solve some of them. Five years later I am happy to say that even though not all these problems have been solved, I am in a somewhat better position to understand them.

The aim of this thesis is two-fold. Firstly, I set out to examine some of the most serious problems that were raised against epistemicism in the literature over the past 30 years. It turns out that the most serious and interesting problems for epistemicism were posed by philosophers generally sympathetic to epistemicist assumptions. My goal was to focus my efforts on these problems. They were largely unanswered by the creator of epistemicism (in its most well-known form), Tim Williamson, so there was something for me to do. Moreover, it seemed to me that solving these problems would be a step to pushing the

[1] MLE - Metaphysics, Logic, Language, and Epistemology. Thanks to Annina Loets, James Kirkpatrick, Alex Roberts, AJ Gilbert, Tomi Francis, Weng Kin San and others for the stimulating discussions during MLE meetings.

theory forward. Not to mention that dealing with such problems was overall a much more stimulating exercise than dealing with the expressions of the incredulous stare that have been a popular response to epistemicism.

The other goal was to deal with semantic plasticity as a phenomenon independent from the epistemicist treatment of vagueness. Recent research indicates that semantic plasticity and epistemicism are not a package deal. The thesis that meanings in our language are shifty is plausible in its own right. I did not manage to focus on semantic plasticity as an independent phenomenon as much as I had wished. Nevertheless, there are parts of the thesis that do this, like my treatment of metalinguistic comparatives in Chapter 8.

1 Vagueness and the Sorites

At the outset it might be a good idea to explain the basic issues regarding vagueness. The topic of vagueness has generated a vast literature. Philosophers disagree as to whether vagueness is best understood as a semantic, epistemic, or metaphysical phenomenon among many other finer details. However, it is generally agreed that vagueness is associated with borderline cases and susceptibility to the Sorites Paradox. Let's start with the notion of a borderline case. Suppose that David is one of those people of whom we would not be able to say whether they are tall. David is not clearly tall; but he is not clearly not tall either. David is an instance of a borderline case of 'tall': it's neither clear that the term applies to him nor clear that it doesn't. A key question that arises, and that is very typical of a borderline case, is: why are we not in a position to know whether David is tall? Before giving an answer to this question, let's consider the Sorites Paradox.

Suppose that there is one grain of sand on the table. One grain does not a heap make: there is no heap on the table. If one grain does not make a heap, adding one more won't make a difference: if there is one more, there is still no heap on the table. However, it seems that if adding one grain more never makes a difference as to whether there is a heap on the table, we can add one grain 9,999 times and we will have 10,000 grains on the table. Then we will clearly have a heap. The argument can be summarised as follows:

(1) 1 grain of sand is not a heap.

(2) If 1 grain is not a heap, then 2 grains are not a heap.

(3) If 2 grains are not a heap, then 3 grains are not a heap.

…

(10,000) If 9,999 grains are not a heap, then 10,000 are not a heap.

(C) 10,000 grains of sand are not a heap.

The argument is valid: we need to apply modus ponens 9,999 times to get from (1), (2), (3)…(10,000) to (C). However, (C) is clearly false: such a large number of grains would make a heap. Since the conclusion is false and the argument is valid, we need to reject at least one premise in the argument.

Epistemicism provides a simple solution to the Sorites. Classical logic, in particular the Law of Excluded Middle (LEM), tells us that for any number of grains of sand they either form a heap or they do not. If 1 grain does not form a heap and 10,000 does, there needs to be some N such that N grains do not form a heap, but N+1 do. The epistemicist takes this claim at face value: one of the conditional premises ((2)-(10,000)) is false. For some N, if there are N grains of sand on the table then adding one more would turn the collection of grains from something that is not a heap to a heap. Thus, for the epistemicist there is a sharp boundary between heaps and non-heaps: there is a smallest number of grains that a heap can have.

Moreover, the epistemicist claims that there is a characteristic kind of ignorance associated with vagueness that prevents us from knowing where that boundary is. The more detailed explanation of this characteristic ignorance is left for the main body of the thesis. Roughly, the epistemicist claims that the reason, why we cannot know where the boundary between heaps and non-heaps lies, is that the meanings of expressions in our language, like 'heap', easily shift in ways undetectable for us. Even if actually some N is the minimum number of grains for a heap, it could easily be the case that it would be some other number (e.g. N+1). We are not perceptive enough to notice the difference.

The reason why the shiftiness of meanings (i.e. *semantic plasticity*) produces ignorance is that it violates the *safety condition on knowledge*. Approximately, a belief is safe if and only if there is no easy possibility of the belief being false. A belief can constitute knowledge only if it's safe. How does semantic plasticity lead to failures of safety? Suppose that the actual boundary for counting as a 'heap' is N grains. However, the meaning of 'heap' could easily be different.

The epistemicist employs the same resources to analyse borderline cases. Take David's case. The epistemicist claims that David is either tall or not tall: there is a sharp boundary between the tall and the non-tall people. Depending on which side David's height falls (whether he is shorter than the shortest tall person), he will be either tall or not tall. The explanation of why we cannot know whether David is tall is the same as in the case of the heap of sand: the meaning of 'tall' could easily be slightly different and David's (actual) height could easily fall on the other side of the boundary. We would not be able to tell the difference. Even if David were actually tall, our belief that he is tall would not be safe. It could easily be the case that the meaning of 'tall' would

be such that David's height would fall on the other side of the boundary. By expressing our belief and uttering 'David is tall' in such a case, we would make an easy error. Since knowledge is incompatible with easy errors (beliefs have to be safe to constitute knowledge), semantic plasticity explains our ignorance in borderline cases.

The epistemicist treats the above cases as examples of inexact knowledge. We are in a position to know some things, e.g. that 1 grain of sand does not make a heap and that 10,000 grains do. However, our knowledge is inexact: it comes with a margin of error. Suppose that the boundary for 'heap' is N grains. Then if m is the margin of error associated with our knowledge of the meaning of 'heap', then we are not in a position to know that the boundary is not N+m. Thus, the epistemicist aim is to use independently motivated epistemic principles to analyse vagueness.

Supervaluationism is the standard foil to epistemicism: the two are arguably the most worked-out accounts of vagueness. The main idea behind supervaluationism is that vagueness is a semantic phenomenon: our use of language does not determine a single sharp boundary. Rather, there are multiple ways to make a vague term precise that would be consistent with (admissible by) our usage. These ways of making vague terms precise, i.e. precisifications or sharpenings, obey classical logic. For instance, on any precisification of the vague term 'tall' there is a sharp boundary between the things that count as 'tall' and those that do not. However, this boundary is in a different place on distinct precisifications. Furthermore, precisifications satisfy the so-called penumbral connections; for instance, if Charlie is shorter

than David, then for any precisification of 'tall', if Charlie classifies as 'tall' then so does David.[2]

The supervaluationist explanation of the Sorites Paradox and the ignorance in borderline cases is slightly more complicated than the epistemicist one. For the epistemicist vague terms simply have sharp, though unknowable, boundaries. The supervaluationist would like to avoid sharp boundaries settled by usage. Rather, the supervaluationist argues that multiple precisifications are consistent with the use of language. Every precisification introduces a sharp boundary for any vague term; however, different precisifications place the boundaries in distinct places. In borderline cases like David's, different admissible precisifications will draw the boundary in distinct places: on some precisifications David counts as 'tall' and on some he does not. Call a sentence supertrue if it is true on every precisification admissible by usage (and superfalse if it is false on every precisification). For instance, if Becky is 200 cm in height, she counts as 'tall' on every admissible precisification. Thus, 'Becky is tall' is supertrue. On the other hand, 'David is tall' is neither supertrue nor superfalse. Another locution that is often used is to say that if some sentence ϕ is supertrue, then 'Definitely ϕ' (denoted also by '$\Delta\phi$') is true.[3]

The (orthodox) supervaluationist equates truth with supertruth and falsity with superfalsity. In borderline cases, when a sentence is neither supertrue nor

[2] The existence of penumbral connections is not a point of contention between epistemicism and supervaluationism as epistemicists accept penumbral connections as well.

[3] The definiteness operator has a dual in ∇ corresponding to the notion of borderlineness. ϕ is borderline if it's neither definitely true nor definitely false: $\nabla\phi \leftrightarrow (\neg\Delta\phi \wedge \neg\Delta\neg\phi)$.

superfalse, it does not get assigned a truth value. Thus, in borderline cases like David's, knowledge is impossible since truth is a prerequisite for knowledge: we are not in a position to know that David is tall, because 'David is tall' is not true. In case of the Sorites Paradox, the supervaluationist argues that there is one among the premises (2)-(10,000) that is false. On every precisification of 'heap', one of the premises is false. However, there is no premise that is false on every precisification. Therefore, the supervaluationist is able to reject the Sorites argument, while at the same seemingly not being committed to there being a definite sharp boundary between heaps and non-heaps.[4]

There are many points of contention between epistemicism and supervaluationism. Here I will briefly mention three. The first point concerns the plausibility of the epistemicist claim that vague expressions in our language have sharp boundaries. The supervaluationist claims that intuitively there are many boundaries that a term like 'heap' could have and be consistent with the way in which we use the term. Our usage is simply not fine-grained enough to pick a particular precisification. The basic epistemicist reply to the challenge is that it is no more plausible that our usage determines a collection of precisifications than that it determines a single one. After all, if our usage is compatible with a collection of precisifications, then it has to be fine-grained enough to draw a sharp boundary between the precisifications belonging to that collection and those that do not. The second epistemicist reply is that sharp boundaries are something that is predicted by our best logical theory;

[4] The epistemicist argument is that this is not a distinguishing feature of supervaluationism. After all, epistemicists also think that there is no definite sharp boundary between heaps and non-heaps, but they interpret 'definite' epistemically. More will be said on this point in Chapter 1.

since we are better than logic than philosophy, we should trust our logical theory and not our prima facie intuitions about what's compatible with usage.

These two replies by the epistemicist illustrate two other main points of contention between supervaluationism and epistemicism. The first concerns higher order vagueness, the second - classical logic.

The problem of higher order vagueness is often thought to be one of the main reasons for preferring epistemicism to supervaluationism. Supervaluationism aims to avoid sharp boundaries for vague terms. Instead of postulating that there is a sharp boundary between heaps and non-heaps, supervaluationism claims that there are multiple precisifications consistent with our usage and no sharp boundary is definitely the one that is selected by usage. However, this raises the problem of higher order boundaries. It seems that the supervaluationist has replaced the sharp boundary between heaps and non-heaps with two other sharp boundaries: the boundary between definite heaps (x's such that 'x is a heap' is supertrue) and borderline heaps as well as the boundary between definite non-heaps (x's such that 'x is a heap' is superfalse) and borderline heaps. This is a problem for the supervaluationist: if the goal is to avoid sharp boundaries, then that goal is not met. Seemingly, all that the supervaluationist has done is to kick the can down the road. Avoiding sharp boundary on the ground level (the boundary between heaps and non-heaps), reintroduces the boundaries at higher levels.

The epistemicist is thought to have an advantage when it comes to dealing with higher order vagueness. Since the epistemicist embraces sharp boundaries on the ground level already, she has no problems to accept them at higher levels. Moreover, the basic explanation of our ignorance is the same for vagueness of all levels: the higher-level boundaries are shifty in the same way that the ground level boundary is and produces ignorance in the same way.

Supervaluationists have made various attempts to solve the problem of higher order vagueness. There is no space to explore them in this crash course on vagueness. Overall, the problem of higher order vagueness is thought to be one of the main points of contention between epistemicism and supervaluationism and many take the simplicity of the epistemicist solution to be its advantage over its rival.

The last point of contention that I will mention here is the issue of logic. The epistemicist claims that, in contrast to supervaluationism, it's able to preserve classical logic. The reason why the preservation of classical logic is important for the epistemicist is that classical logic is used by mathematicians. The success of mathematics as a discipline suggests that we should not revise the logic they use to provide an account of vagueness (especially that there are theories like epistemicism that don't require such revisions). What does it take for logic to count as classical according to the epistemicist? Firstly, it has to validate all the classical tautologies like the Law of Excluded Middle (for any ϕ, $\phi \vee \neg\phi$ is true) and not validate any classical contradictions (like sentences of the form $\phi \wedge \neg\phi$). However, all the classical tautologies are validated by supervaluationism: all the classical tautologies are true on every precisification, and all the classical contradictions are false on every precisification. Where the epistemicist and supervaluationistic logic come apart is the meta-rules of proof.

The supervaluationist equates truth with supertruth. If we understand logical consequence as truth preservation in every model, then on the supervaluationistic understanding truth as supertruth, ψ is a consequence of ϕ if and only if ϕ is supertrue whenever ψ is supertrue.[5] It turns out that on such

[5] This is the global definition of validity. On the local definition of validity, ψ is a consequence of ϕ if and only if on any precisification, if ϕ is true then so is ψ. The problem with the local validity is that it does not make use of the notion of

an account of logical consequence, some meta-rules of proof employed by mathematicians break down.[6] One of such rules is proof by cases.[7] Suppose that we have a disjunctive premise A∨B and we are proving that C follows from our premise. The standard way of proving this is to conduct a proof by cases: prove that C follows from A and prove that C follows from B, to conclude that C follows from A∨B. However, supervaluationists have to reject this rule. $\Delta\phi \vee \Delta\neg\phi$ is a logical consequence of ϕ as well as a logical consequence of $\neg\phi$; nevertheless, it cannot be the case that $\Delta\phi \vee \Delta\neg\phi$ is a consequence of $\phi \vee \neg\phi$, because this would mean that we are not allowing for the possibility of borderline cases (when neither $\Delta\phi$ nor $\Delta\neg\phi$ is true). The inference from $\phi \vee \neg\phi$ to $\Delta\phi \vee \Delta\neg\phi$ is not valid; thus, it seems that we have to reject the meta-rule of the argument by cases. For this reason, the epistemicist claims that supervaluationistic logic is revisionary relative to classical logic (we might call the supervaluationistic logic semi-classical).

The standard supervaluationist response is to say that in fact the revisions made to logic in no way affects the way that mathematic is done. The reason is that mathematicians use precise terminology in their proofs.[8] Mathematical notions (such as 'set', 'function' etc.) are precisely defined and have no borderline cases anyway. If we only use precise vocabulary in our proofs, then we can still use the classical meta-rules for proof, because if ϕ is precise then it is equivalent to $\Delta\phi$. Of course the epistemicist will object that not all the

---

supertruth. However, the thought is, if the supervaluationists equate truth with supertruth, then supertruth should also feature in the definition of logical consequence. See Williamson (1994b, p. 148-149) on this point.

[6] Although see Williams (2008b) for a dissenting view.

[7] Other cases include contraposition, conditional proof and reductio ad absurdum. See Williamson (1994b, p. 151-152).

[8] See McGee & McLaughlin (2004).

notions that we use mathematics (and mathematical techniques) for will be precise. That might be true for pure mathematics, but not will not be in general true for applied mathematics: sometimes we want to apply mathematics to real world notions that are not as precise as the language of pure maths.

I do not intend to resolve these controversies here. My aim was merely to present the basic notions used by two of the most prominent theories of vagueness and outline some of the major points of contention between them. Of course, there are many more options other than epistemicism and supervaluationism when it comes to explaining vagueness. However, I will not focus as much on these other views in the thesis, so my crash course introduction to vagueness covers only these two rivals.

2 How to read the thesis

The thesis consists of eight chapters. Chapter 1 sums up the debate over epistemicism that has emerged after the publication of *Vagueness* by Timothy Williamson in 1994. I try to identify the main points of contention between epistemicism and its critics. In some cases, I consider a debate to be closed: usually this is because Williamson (in my assessment) has done a sufficiently good job in answering the critics and I have nothing more to add on that point. In other cases, I believe that the epistemicist can say more to defend and clarify their position. In such cases, debates are introduced briefly and are dealt with in greater detail in subsequent chapters.

Part I of the thesis, which follows the introductory Chapter 1, is concerned with metasemantic matters. Chapter 2 presents the problems emerging from the debate over the metasemantic aspects of the epistemicist theory and sets out to answer them. Chapter 3 attempts to answer a particularly difficult

problem for epistemicist metasemantics posed by Moral Twin Earth cases and objections by Miriam Schoenfield (2016). I present an epistemicist treatment of moral vagueness that avoids these problems.

Part II of the thesis is concerned with more technical issues. Chapter 4 is concerned with factor relative safety. The epistemicist understands ignorance in borderline cases as a failure of safety due to a particular factor: semantic plasticity. However, it's unclear how to give a rigorous treatment of such claims. I attempt to present a general analysis of safety relative to a factor. Chapter 5 develops an account of the definiteness operator for the epistemicist theory. Epistemicism is a well-understood theory, but the epistemic interpretation of 'definitely' has been surprisingly difficult to capture. I present an epistemicist treatment of definiteness that avoids some pitfalls that other accounts face. Chapter 6 attempts to answer the question whether metalinguistic expressions such as 'truth' are vague and considers the implications for the semantics of such vocabulary in the light of the account of definiteness developed in Chapter 5. For this reason, I would recommend reading Chapters 5 and 6 together.

In Part III, I focus on selected puzzles related to epistemicism and semantic plasticity. Chapter 7 attempts to solve a puzzle for epistemicism due to Adam Sennet (2012). Sennet argues that an omniscient speaker could prevent a term such as 'rich' from being semantically plastic without consequently preventing it from being vague, which would call the epistemicist explanation of vagueness into question. Chapter 8 develops a novel account of the semantics of metalinguistic comparatives: sentences like 'his problems are more financial than legal' or 'Gödel was a logician more than a philosopher'. The account uses the notion of similarity of interpretations borrowed from the epistemicist

treatment of semantic plasticity. Chapters in Part III can be read independently from the rest of the thesis and of one another.

Chapter 1

Epistemicism: 30 years after

ABSTRACT

In this chapter I introduce Williamsonian epistemicism and outline the main points that were discussed in the literature in the years following the publication of *Vagueness* by Timothy Williamson.

WORD COUNT: 11,400

1 Introduction

Epistemicism is one of the most systematically developed theories of vagueness. Systematicity is a double-edged sword. As is usually the case, when a theory is spelled out precisely it is easier for opponents of the theory to come up with arguments against specific aspects of the theory that go wrong. Thus, it is unsurprising that the epistemicist theory has been criticized from multiple angles. In this chapter, I will focus on outlining the theoretical commitments of epistemicism in the version defended by Timothy Williamson (1994b) and the major lines of criticism of the epistemic view of vagueness. This should also make it clear what the starting point of the project is and what question the thesis should aim to answer.

The critique of the epistemicist theory of vagueness has proceeded in roughly three stages. The first stage was focused on the epistemicist metasemantics, i.e. the part of the theory that explains how expressions in our language are assigned the meanings that they do.[9] Critiques of this kind are usually expressions of the famous 'incredulous stare' directed at the epistemicist. The general idea is that even if epistemicism is a neat theory, it simply must be false, because the consequences of the theory in metasemantics are simply so implausible that they outweigh the potential benefits of endorsing epistemicism.

The second stage has been more sympathetic to epistemicism.[10] It proceeds by accepting many controversial theses that are endorsed by epistemicism and attempting to show some problems from 'within' the theory. These criticisms are mainly aimed at the models of margins for error, the safety principle, and the epistemicist account of borderline cases (in particular the semantics of the definiteness operator). This group of arguments is more technical and more challenging for the epistemicist, because they are made from positions that are not generally hostile to epistemicism. Many of the problems raised for epistemicism that belong to the second stage criticisms have not been answered in the literature; the bulk of this thesis will be devoted to moving the epistemicist project forward by providing answers to these problems.

The third stage attempts to undermine the status of epistemicism as the unique theory that can retain classical logic and classical semantics. Epistemicism owes a lot of persuasive force to the fact that it provides an

---

[9] See Burgess (2001), Gómez-Torrente (1997), Horgan (1997), Horwich (1997), Keefe (2000), Machina & Deutsch (2002), McGee & McLaughlin (1995; 1998; 2004), Schiffer (1997; 1999), Tye (1997), Weatherson (2003), Wright (1995).

[10] See Bacon (2018), Caie (2012), Dorr & Hawthorne (2014), Gomez-Torrente (2002), Graff Fara (2002), Hawthorne (2006), Kearns & Magidor (2008), Magidor (2018), Sennet (2012), Yli-Vakkuri (2016).

explanation to the key phenomena associated with vagueness while retaining classical logic and semantics. If it is not the only theory that is able to retain them, then there is less reason to be an epistemicist. Some philosophers argue that on a balance of costs and benefits, alternatives to epistemicism that retain classical logic and semantics do better than epistemicism.[11]

The thesis focuses on the first and second stage of the debate on epistemicism. My aim in this chapter is to outline this debate. Some of the points of contention mentioned here are explored in other chapters. In such a case I try to only flag them and leave the proper discussion for later. In other cases, when I don't discuss a particular point later in the thesis, I try to say something more on that point in this chapter.

In the following sections I will present (a) the outline of epistemicist theory as defended by Timothy Williamson and (b) the debate on epistemicism in the aforementioned aspects. My aim will be two-fold. Firstly, it will be useful to know what exactly the Williamsonian epistemicist is committed to. Secondly, I will try to outline the debate about epistemicism and assess which are indeed serious problems for epistemicism and which ones can be met with a quick rebuttal.

2 Summary of the Epistemic View of Vagueness

The strength of epistemicism consists partly of the fact that it provides an explanation of the problems associated with vagueness by using established philosophical theories and principles. Williamson's (1994b) argument for epistemicism proceeds in three stages. Firstly, Williamson argues that vague expressions have sharp boundaries. Secondly, he argues that if they have sharp

[11] See McGee & McLaughlin (1995), Barnes (2010), Cameron (2010), Barnes & Williams (2008a; 2011), Bacon (2018).

boundaries, then these boundaries are bound to be volatile: they shift across the space of nearby possible worlds. Thirdly, Williamson argues that we can use well-established epistemological principles, in particular the safety condition on knowledge, to show that the shiftiness of boundaries of vague expressions is bound to produce ignorance that we associate with borderline cases. Below I briefly outline epistemicism.

The starting point is classical logic. Williamson argues that this is the default logic that we start from; his aim is to show that we can provide an account of vagueness that is compatible with this default setting. Williamson (1994b) shows that classical logic and intuitive principles about truth entail bivalence. Absent additional ideology, bivalence guarantees that vague expressions have sharp cut-off points, i.e. bivalence entails that for any vague predicate *F* and for any *x*, '*F(x)*' is either true or false.

Consequently, Williamson argues that if vague expressions have sharp boundaries, then these boundaries are shifty.[12] To support this Williamson presents a toy metasemantic model. What words in our language mean has something to do with the way that they are used. Suppose then that the meanings of expressions are determined by the pattern of their use in a linguistic community. Furthermore, our use of language could easily be different. It is natural to suppose that slight changes in use across close possible worlds produce slight changes in meaning. We are insensitive to slight shifts in meaning. Our knowledge of the meanings of expressions in our language is inexact, we don't notice when the meaning changes slightly. Furthermore, we don't have access to all the facts about the use of language in the community nor do we know how these facts about use determine the meanings of expressions. Thus, vague expressions are semantically plastic: intensions of

[12] Williamson's arguments for this are explored in greater detail in Chapter 2.

these expressions could easily be slightly different due to small changes in use that we could not detect.[13]

Williamson (1994b) argues that given semantic plasticity we can use recognised principles in epistemology, namely the safety condition on knowledge, to explain ignorance in borderline cases. Knowledge requires safety from error: if a belief is to constitute knowledge, it could not easily be the case that the belief is wrong. Semantic plasticity can produce such easy errors in borderline cases. Even if a borderline belief expressed using a sentence ϕ is actually true, ϕ could easily be assigned a different intension in which case if one were to express their belief using ϕ, one would say something false.[14] Thus, semantic plasticity is an ignorance inducing factor in borderline cases.

Almost every step of Williamson's argument has been criticized in the literature. The argument can be summarised as follows:

(1) Classicism	Classical logic is the one true logic.
(2) Bivalence (*from (1)*)	Truth is bivalent: if an utterance u says that p, then u is either true or false.

[13] I express semantic plasticity as a variation in the intension across nearby worlds. However, depending on how detailed our analysis is, this may not be the right way of stating the epistemicist commitments. For instance, if we consider a language which includes indexicals and we embrace the orthodox Kaplanian approach to indexicals, then semantic plasticity should not be stated in terms of the variation in intensions (a functions from worlds to extensions), but rather in terms of the variation in characters (functions from worlds to intensions). In chapter 5, I develop an account of definiteness that employs a 4D model. On this model semantic plasticity is cashed out as a variation in the metasemantic character (a function from worlds to characters) assigned to expressions. For most purposes the talk about semantic plasticity as variation in intensions is fine and I will continue to do so for the sake of simplicity, but it's worth pointing out that this might be a simplification.

[14] This implies that it's possible for a belief to have different contents at different worlds. If one does not like this consequence, we can redescribe the situation using a notion of a 'counterpart belief'.

(3)	Sharpism (*from (2)*)	Vague predicates have sharp cut-off points.
(4)	Supervenience	Meanings supervene on use.
(5)	Metasemantic Thesis (*from (3) & (4)*)	Sharp cut-off points supervene on use.
(6)	Safety	Knowledge requires safety from easy error.
(7)	Use Shifts	Small shifts in use of vague terms are easily possible.
(8)	Intension Shifts	Small shifts in use of vague terms can cause small shifts in intensions.
(9)	Indiscrimination	Small shifts in intensions of vague terms are undetectable.
(10)	Semantic Plasticity (*from (5), (7), (8), (9)*)	Intensions of expressions of vague terms vary undetectably across close possible worlds.
(11)	Epistemicism (*from (6) & (10)*)	Semantic plasticity makes our beliefs in borderline cases unsafe; this explains ignorance due to vagueness.

These are the key commitments of the epistemic theory. I will attempt to explicate them in further parts of the chapter and outline the debate in the literature about them.

At the outset it is worth clarifying what is meant by 'use' here. It seems that we could construe use very narrowly, so that the notion encompasses something like the pattern of assent and dissent to utterances within a linguistic community. On such a construal, Use Shifts are very plausible: it

could easily be the case that the pattern of assent and dissent would be different. However, on the narrow understanding of use, Supervenience becomes implausible. For instance, it is implausible that the meaning of natural kind terms like 'water' supervenes on patterns of assent and dissent as shown by Putnam (1975). So use will have to be construed rather broadly, e.g. to include the properties that causally regulate the patterns of assent and dissent within the community. Therefore, we should take 'usage' (as used by the epistemicist) to be construed broadly and serve more as a catch-all term. Broadly construed usage includes narrowly construed usage (patterns of assent and dissent). Thus, a change in narrowly construed usage implies a change in the broadly construed usage, especially in case of standard vague terms like 'tall' or 'bald' (for which it's plausible that the only salient meaning-determining factor is the pattern of assent and dissent). Therefore, we should have no problems running the argument through using broadly construed usage.

In the further parts of the chapter, I mean to outline some of the main points that were raised in the literature on epistemicism. I do not mean to provide a fully comprehensive and conclusive outline of the debates. The literature on epistemicism is too vast to achieve this in a scope of a single (though long) chapter. Rather, my intention is to gesture at the general lines of critique of epistemicism as well as some of the responses to said critiques that emerged in the literature. My hope is that this chapter will serve as an introduction to the debate over epistemicism that will enable me to describe how different parts of this thesis fit with the rest of the literature.

3 Logical Issues

3.1 Classicism

One of the main commitments of epistemicism is its alliance to classical logic. The epistemicist aim is to show that if we are independently committed to classical logic (or at least we have an initial reason to prefer it), then problems pertaining to vagueness can be solved without giving up classical logic. I will use the term 'classical logic' to refer to a system of logic that includes all classical tautologies as well as classical rules and meta-rules (e.g. proof by cases or reductio). 'Classical semantics', as I use the term, refer to truth-conditional semantics based on a bivalent notion of truth. It is often a matter of dispute whether supervaluationistic logic, which accepts all classical tautologies, but does not endorse some classical meta-rules (reductio, proof by cases, deduction theorem) is to be called 'classical'. As stated in the thesis introduction, I will use the term 'semi-classical' to refer to the logic endorsed by supervaluationists and retain the term 'classical logic' for the logic used by mathematicians.[15]

The epistemicist claims to have an advantage over most other theories: she is not forced to provide a revisionary logical and semantic theory but can retain the full power of classical logic and classical semantics. The dialectical position of the epistemicist is that there is an initial bias in favour of classical logic; the epistemicist claims that vagueness does not give us a reason to remove that initial bias, so classical logic is the one that we are left with. However, there may be a dialectical worry about the epistemicist approach: why start from classical logic in the first place?

[15] See McGee & McLaughlin (1998; 2004) and Williamson (2004) for an exchange on this point. See also Williams (2008b).

To give a fully satisfactory answer to the question, the epistemicist would have to argue for something like the thesis that classical logic is the 'one true logic'. At least it would have to be argued that we are justified in being partisan in favour of classical logic.[16] I will not be able to provide a full answer to these problems – their scope would require a whole project of their own.

Nevertheless, we can provide a sketch of why the epistemicist would treat classical logic as a good starting point without claiming to provide decisive arguments for that position. The epistemicist commitment to classical logic does not have to be treated as dogmatic. It is not the case that epistemicists such as Williamson think that the adequacy of classical logic is a Moorean fact that is beyond dispute (Williamson 2002b). There may be reasons to revise some parts of classical logic, e.g. the law of excluded middle, if there is sufficient evidence that classical logic cannot appropriately deal with certain phenomena. The epistemicist position is that vagueness is not such a phenomenon: vagueness can be explained without giving up classical logic.

Why should classical logic be treated as a default? One important reason for the epistemicist to treat classical logic as default is that it is used by mathematicians: the success of mathematics speaks to the appropriateness of the tools they use. This argument was discussed briefly in the introduction to the thesis. Another reason that the epistemicists could give us for treating classical logic as default is to argue for classical logic on the basis that it stands out as a logical theory on some general grounds that are used in evaluating scientific theories, such as simplicity and strength. Such a line of argumentation would require taking the anti-exceptionalist position about logic.[17] On the anti-exceptionalist stance, logic is continuous with science: logical theories can be

[16] See Woods (2019) for a discussion of the problems in evaluating logical theories.
[17] See Quine (1951), Williamson (2013; 2015), Priest (2006), Russell (2014).

evaluated using the same standards as scientific theories. Factors such as explanatory power and simplicity would then count towards evaluation of a theory.[18] The illustration that classical logic is the winner when evaluated on anti-exceptionalist grounds is too great a task to undertake here. However, this is another route that the epistemicist can undertake to defend their treatment of classical logic as default.

I will treat classical logic as a natural starting point to the discussion. There is no doubt that a commitment to classical logic is one of the key commitments of epistemicism.

3.2 Bivalence and Disquotational Principles for Truth

The argument in favour of bivalent semantics presented by Williamson (1994b) is based on the disquotational principles for truth. There is disagreement about the nature of the truth predicate and its semantics, but there are generalisations that we can make about *truth* that anyone can agree on. One such generalization is that an utterance is true if and only if what the utterance says is the case is the case. In other words:

(Truth) If an utterance u says that p, then u is true if and only if p.

Similar principle holds for the notion of falsity. An utterance is false if what it says is the case is not the case. Thus, the following principle captures the basic features of falsity:

(Falsity) If an utterance u says that p, then u is false if and only if not p.

[18] Although it should be noted that there are many issues with evaluating logical theories on such general grounds. See Woods (2019).

The first thing to note that these principles are not instances of the standard Tarskian biconditional of the form:

'P' is true if and only if P.

For one thing, instances of the Tarskian schema are contingent: had 'P' expressed some different proposition, the biconditional could be false. On the other hand, Truth is necessary: necessarily if u says that p, then u is true if and only if p. Another difference is that the Tarskian schema is context sensitive. When uttering an instance of the Tarskian biconditional we assume that the context relative to which we evaluate both sides of the biconditional is the same for both. Furthermore, we can't easily get rid of this context dependence by quantifying over all contexts, e.g. it might not be the case that 'I am tall' is true in context c if and only if I am tall in context c, when in context c the speaker is short and is not me. No such problems with context-dependence arise for Truth. Lastly, some argue that the T-Schema should be restricted because of the Liar Paradox, e.g. because the Liar sentence fails to express a proposition (or say anything). Such a position does not entail a restriction on Truth (because the antecedent of Truth is false in case a sentence/utterance fails to express a proposition, which makes the whole conditional true).

Therefore, the most popular qualms about the T-Schema are not relevant to the truth of Truth. Even theorists that don't endorse the disquotational theory of truth have to account for the disquotational nature of truth: it is part of the role of the truth predicate to move from the metalanguage to the object language and to disquote sentences of the object language. Truth and Falsity are very plausible candidates to capture this role of truth. However, these principles, along with classical logic, entail bivalence. Suppose that u says that p. We know from classical logic that the law of excluded middle (LEM)

holds. If LEM holds, then so does its instance. Therefore, p or not p. If p, then u is true (by Truth). Therefore, if p, then u is true or false. If not p, then u is false (by Falsity). Therefore, if not p, then u is true or false. Thus, u is true or false. Discharging the hypothesis, we get:

> (Bivalence) If an utterance u says that p, then u is true (if p) or false (if not p).

Historically the debate over whether truth obeys bivalence has been treated as a key issue in the debate between epistemicism and supervaluationism. The reason for this is clear: traditional supervaluationists (Fine 1975; Keefe 2000) equate truth with supertruth (truth on all admissible precisifications of the language). Supertruth does not obey bivalence but admits truth-value gaps instead. Establishing that intuitive disquotational principles and classical logic entail bivalence was an important step in that debate. However, the issue is not as simple as it seems at first glance.

Andrew Bacon (2018) argues that the debate between epistemicists and supervaluationists regarding bivalence is insubstantial.[19] That is because supervaluationists may also add a bivalent predicate to the language: there is nothing in their theory that would make the notion incoherent. They may even choose to call the bivalent predicate 'truth' and reserve the term 'determinate truth' for supertruth. That is precisely the option chosen by supervaluationists such as Vann McGee and Brian McLaughlin (1995; 1998; 2004). Introducing the bivalent notion of truth allegedly allows the supervaluationist to adopt the terminology of the epistemicist and escape the force of the argument from bivalence.

[19] A similar position is expressed by Schiffer (1997).

I think that the claim that the debate is insubstantial is too strong. There is something that the bivalent supervaluationists such as McGee and McLaughlin (1995) accept that the traditional supervaluationists reject, namely that the bivalent notion of truth plays some part of the role that we attribute to truth. For instance, the bivalent notion of truth may play a role in our reasoning and allow us to make inferences that we would not be able to make with the supertruth notion of truth (such as an inference from P to 'P is true', when P is borderline). This is a fairly obvious point: why introduce a notion of truth if it is not to play any role in out theorising. Therefore, the introduction of the bivalent truth predicate, if it is to soften the blow of the argument from bivalence, means that likely one must accept that bivalent truth plays a role in our theorising, which constitutes a departure from standard supervaluationism.

Is this departure something that is desirable from the supervaluationist perspective? I think that it is, given the problem posed by higher-order vagueness. The alleged problem with the bivalent notion of truth is that it has borderline cases: it's borderline whether 'P' is disquotationally true whenever it is borderline whether P; such a notion of 'truth' seems to be unsuitable for objective semantics. However, we can recognise that supertruth is plagued with the same problem as disquotational truth once we move to second-order vagueness: it is vague what interpretations count as 'admissible', so it is vague what cases are supertrue (Williamson 1994b, p. 163). Therefore, the alleged advantage of supertruth disappears. However, there are significant theoretical costs to equating truth with supertruth; in particular, such a position requires the rejection of the validity of the disquotational principles and the Tarskian T-Schema. The supervaluationist is better-off endorsing the bivalent notion of truth.

However, some critics remain unconvinced.[20] The thought is that it seems implausible that there is a sharp cut-off point between the tall and non-tall people. Suppose that a person that is N millimetres in height is not tall and a person that is N+1 millimetres in height is tall. What makes it the case that the latter is, and the former is not tall? What is the fact that makes it so?

The epistemicist can say that the former person is not tall because of the fact that they are not tall and that the latter person is tall because of the fact that they are tall. If opponents of epistemicism think that these facts don't count as *real* facts, they have to explain why. However, even if the epistemicist is technically correct, such a reply is not very informative. After all, 'it is the case that x is F because of the fact that x is F' sounds like truism. A better epistemicist reply invokes supervenience principles that are accepted by the critics. Presumably, the fact concerning whether one is tall or not supervenes on their height. If a person's height is sufficiently great, they are tall; otherwise they are not tall. A person who has N mm in height is not tall because of their height: it's too small to qualify them as being tall; the person whose height is N+1 mm is tall because their height is great enough for them to have the property of being tall.

The crucial point is that according to the epistemicist, any vague property such as 'being tall' is metaphysically equivalent (necessarily coextensive) to a precise property. Since the opponents of epistemicism see no problem in explaining how objects can instantiate precise properties and how it is possible for such properties to have sharp cut-off points, they should see no problem with the epistemicist explanation of how objects instantiate properties that are metaphysically equivalent to the precise ones. At this point the debate usually moves from the level of the world to the level of the language. Maybe the

[20] E.g. see Burgess (2001).

epistemicist reply works for explaining why a person with N mm in height is not tall and a person with N+1 mm in height is tall, but what is the explanation for why the former does not fall in the extension of 'tall' and the latter does? Fortunately for the epistemicist the same answer can be given to such a question: a person with N mm in height is too short to be in the extension of 'tall' and the person with N+1 mm in height is not. So it seems that the opponent of sharp cut-offs cannot point to a problem with the epistemicist story so far. I will come back to this point in Chapter 2.

I agree with Bacon (2018) that once the supervaluationists accept that bivalent truth plays the theoretical role that I described, the debate regarding truth between supervaluationists and epistemicists will become somewhat insubstantial.[21] For example, supervaluationists will claim that determinate truth still plays an important theoretical role that is not played by bivalent truth, e.g. being the norm for assertion: p may be asserted only if it's determinately p. However, this is not something that is challenged by the epistemicist. The epistemicist will simply interpret the talk of 'determinacy' epistemically.[22] Both parties would agree that indeterminacy entails unknowability and that knowability entails determinacy. The question that remains is whether there is a notion of 'determinacy' that is not fully captured by the epistemicist notions.

An important problem is dissolved once we make the above observations. In particular, some opposition regarding Sharpism can be dismissed. Epistemicists claim that the crucial consequence of Bivalence is Sharpism:

---

[21] Williamson (1994b; 2004) argues that once supervaluationists make this move their position is indistinguishable (apart from the rhetoric) from the epistemicist position.

[22] Chapters 4 and 5 explore the epistemicist treatment of determinacy (definiteness). Some epistemicists, e.g. Williamson, propose an even stronger norm of assertion: p may be asserted only if it is known that p. However, such a norm is not required if one's goal is to mimic the supervaluationist principle.

(Sharpism) Vague predicates have sharp cut-off points.

Note that by the classical law of excluded middle for any vague predicate F and for any x, it is either the case that Fx or it is the case that $\neg Fx$. For instance, by the least number principle[23], classical logic guarantees that there is a shortest tall man, the poorest rich man, and the smallest heap of sand. Bivalence additionally guarantees that for any vague predicate F and any x, if 'Fx' says that p, then 'Fx' is either true (if p) or false (if not p). This means that there is a shortest man for whom the predicate 'tall' is true of, i.e. that the predicate 'tall' has a sharp cut-off point.

Some could worry that the epistemicists claim that there is a determinate cut-off point for every vague predicate. But that is a misreading of the epistemicist theory. The epistemicist is not saying that we can know where the cut-offs for vague predicates are. They claim that it is unknowable due to vagueness where the cut-offs for vague predicates are. Thus, the actual cut-off point is not a definite cut-off point. Nevertheless, it's definitely the case that there is a cut-off point according to the epistemicist.

However, this is not something that the supervaluationist may plausibly deny. According to supervaluationism, any admissible precisification is classical. Consequently, it is determinately the case that there is a shortest tall man, and it is determinately the case that there is a shortest man of whom 'tall' is true of (as this is the case on any admissible precisification). So supervaluationism is also committed to sharp cut-off points in the same way that epistemicism is. Therefore, both sides of the debate agree that vague predicates definitely have sharp cut-off points and that no cut-off point is definitely the actual cut-off point. If one wants to attack epistemicism by

[23] The least number principle states that any non-empty set of natural numbers contains a least element.

challenging Sharpism or the inference from Bivalence to Sharpism, the point has to be more subtle. I will outline such criticisms in Section 4.

3.2.1 Definite Truth

The epistemicist is committed to understanding 'definiteness' and 'determinacy' as epistemic notions (when they are used in contexts pertaining to vagueness).[24] An utterance of 'Tom is definitely rich' means something like 'Tom is knowably rich' on the epistemicist reading. This is not exactly right because the epistemicist is not committed to the view that every definite truth is knowable: there may be reasons for unknowability that have nothing to do with vagueness. So 'definitely p' means something like 'there are no obstacles characteristic of vagueness to knowing that p'.

Opponents of epistemicism will argue that there is a sense of 'indefiniteness' or 'indeterminacy' that is non-epistemic. However, the question is whether we would ever need such a notion for theorising. According to the epistemicist, the explanation of key issues associated with vagueness does not require non-epistemic notions of definiteness. If there is indeed non-epistemic indeterminacy (e.g. ontic indeterminacy[25]) the epistemicist can take a 'quietist' position about it. We don't need to invoke the notion to explain issues related to vagueness. However, it can be an interesting subject for deliberation if the notion turns out to be coherent and useful about theorising about other matters.

---

[24] The epistemicist is not committed to the epistemic understanding of 'determinacy' and 'definiteness' in other contexts, for in instance epistemicism is compatible with the view that the future is metaphysically indeterminate.

[25] See Barnes (2010), Barnes & Williams (2011), Cameron (2010).

4 Metasemantic Issues

A large part of the debate about epistemicism in the 1990s after the publication of *Vagueness* has been concerned with metasemantic issues. These will be explored in greater detail in Chapter 2 (and to some extent in Chapters 3 and 6). The general strategy of the critics of epistemicism has been to argue that epistemicism requires an implausible metasemantic theory. The famous incredulous stare 'challenge' to epistemicism is usually directed at this aspect of the theory. There are broadly two points of criticism that are packed into the incredulous stare. They are summed up by the following questions:

(1) How could vague expressions have sharp cut-off points?

(2) How could it be the case that our use selects a particular cut-off point and not some other that seems equally good?

I will present them very briefly below. Of course, these challenges have been stated many times in subtly different ways, so the way I present them is not the only possible one. At the end of Section 4, I will also a challenge to the claim that meanings of expressions in our language are shifty in the way required by epistemicism.

4.1 Incredulous Stare 1 - The Sceptical Challenge

McGee & McLaughlin (2004) argue that what is incredible in the epistemicist story is that it relies on an unfounded metasemantic generalisation about our use of vague expressions. Epistemicism does not require only that there are sharp cut-off points, but also relies on the model on which the sharp cut-off points are determined by use (alongside some *semantic laws* that connect usage with meaning). The determination of cut-off points on use is usually cashed out in terms of supervenience:

(i) meaning supervenes on use

(ii) vague expressions have sharp cut-off points

(iii) the cut-off of an expression supervenes on the expression's meaning

(iv) by transitivity of supervenience, cut-offs supervene on use.

McGee & McLaughlin (2004) find it incredible that our meaning could determine a particular cut-off point. One way to present the worry is to question (iii): perhaps vague expressions have sharp cut-off points, but these do not supervene on the expression's meaning. However, rejecting (iii) requires denying the highly intuitive principle according to which the extension of an expression supervenes on its meaning; in other words, no change in the extension on the term without a change in meaning.[26] So it might not be plausible to challenge the supervenience theses. The problem is that epistemicism seems to need a stronger notion than supervenience for its purposes. In particular, epistemicism requires that our use is fine-grained enough that there is only one cut-off point that fits best the pattern of use.[27] So the epistemicist relies on the metasemantic generalisation that our pattern of use of the language is such that for any vague expression there is a sharp boundary that best fits that use.

Epistemicism is not unique in this regard. For example, supervaluationism relies on the metasemantic generalisation that there are many, and not just one, potential cut-off points in case of vague expressions that fit the pattern of use equally well. However, some philosophers[28] (a) take the supervaluationist generalisation to be intuitively more plausible or (b) think that we should wait for further developments in metasemantics to decide which theory is correct.

---

[26] I am omitting any issues that context-sensitivity might pose here.

[27] This could be also phrased in terms of consistency with use, however the speakers' use could be conflicting; in such a case, there would be no candidate that would be consistent with everyone's use, but there still could be one that best fits the pattern of use.

[28] See McGee & McLaughlin (2004).

The challenge to epistemicism is to explain how our metasemantic theory would explain the claim that our use determines sharp cut-off points for vague expressions. What metasemantic law says that vague expressions must be sharply bounded? McGee and McLaughlin (2004) argue that this problem is especially pressing for Williamson because he is sceptical that we would be ever able to learn enough about metasemantics to learn what the cut-off points for our vague expressions are. With such a sceptical view of the prospects for the development of metasemantics, one should not base one's theory on grand metasemantic generalisations such as that our use fixes sharp boundaries for vague expressions. This objection will be explored in greater detail in Chapter 2.

4.2 Incredulous Stare 2 - which cut-off?

A related objection to the epistemicist theory states that epistemicism cannot explain why a particular cut-off point is selected by use. Epistemicism requires that every vague predicate draw a sharp boundary between the things in its extension and the things in its anti-extension. For example, for some number N, N is the smallest number of dollars that one can have and be rich; one cent less, and one is not rich. Thus, the boundary between things that fall in the extension of 'rich' and those that fall in the anti-extension of 'rich' is drawn at N. Why should we take it that N is the cut-off point for 'rich'? Supervenience and Sharpism together entail that there is a cut-off, but they don't tell us where it is or what makes it the case that this particular cut-off is picked out. There are two versions of this challenge: one regards the vagueness of metalinguistic terms such as 'truth' (Keefe 2000, Chapter 3); the other is concerned with the very real possibility that there are multiple candidates for being the cut-off point of a vague expression that are on a par with regards

how they fit the use (Burgess 2001; Weatherson 2003). The former challenge is addressed in Chapter 6; the latter in Chapter 2.

4.3 Intension Shifts

Williamson is committed to small shifts in use of vague expressions causing small shifts in their intensions; this is required for the explanation of ignorance due to vagueness via semantic plasticity to go through. For the epistemicist, it cannot be the case that only large shifts in use cause shifts in intension because worlds where we use words in a radically different way do not count as relevantly close; if that were the case, then the Safety Principle could not be used to explain ignorance due to vagueness.

Several authors expressed doubts about the thesis that vague words are as shifty in their reference as Williamson's theory suggests.[29] Perhaps it is the case that once the boundary for a vague term is fixed, it does not change due to small shifts in use, but only when the shift in use is quite large. Williamson has not produced evidence that small shifts in use are sufficient to shift the reference of vague expressions.

Williamson (1997a) argues that Intension Shifts is not an indubitable thesis, but that it is simply more plausible than the alternative.[30] If the reference would shift only after large shifts in use, then the actual cut-off and

[29] See Keefe (2000, Chapter 3), Horwich (1997), Mahtani (2004), and Schiffer (1997; 1999).

[30] Cf. Mahtani (2004), who argues that the alternative picture is more plausible because the only way that words could have sharp meanings is if they have a natural referent. Cameron (2010) also argues that sharp cut-off points are due to some candidates for reference scoring higher on naturalness ranking; however, according to Cameron it's highly contingent which properties are more natural than others, which makes his view compatible with shifts in reference across close possible worlds (the difference being that the shifts in reference are produced by the shifts in naturalness ranking and not usage facts).

the nearest point where the reference shifts would mark natural boundaries of sorts. But the standard presupposition in borderline cases is that none of the admissible candidates for the sharp boundary of a given expression is more natural than others. So Williamson's picture is more plausible than the alternative.

It's possible to push back on Williamson's reply by noting that it is quite likely that meanings in our language can exhibit a kind of 'stickiness'. Once a reference has been fixed when a given term is introduced into the language, the reference does not change unless there is a significant change in use. Consider the word 'meter'. Its reference has been introduced by reference to a particular piece of metal. It is obvious that we could have used a piece of metal of different length that would be equally well suited to do the job. But we chose the one that we did and now the reference of the word 'meter' is fixed and slight variations in use do not change it. Perhaps this is also the case with other meanings: once the reference of the word 'tall' has been initially fixed, it does not change unless there is a significant shift in use.

Such a picture may be plausible, but it is toothless against the epistemicist. For the epistemicist does not need it to be the case that there are shifts in the reference of a vague word across time in the actual world. It may be very plausible to suppose that vague words are both modally plastic (variance across close possible worlds) and temporally plastic (variance across time). However, only modal plasticity is essential to epistemicism, because it is required for the safety-based explanation of ignorance due to vagueness to go through. The picture sketched above does not go against modal plasticity; even if meanings of words are sticky to their initial meaning (like 'meter') this does not mean that they are not modally plastic. The epistemicist account of reference shifts is still the most plausible one.

5 Epistemic Issues

An important group of challenges to epistemicism is directed at the epistemology underlying the epistemic view of vagueness. The challenge for the epistemicist is to explain why vagueness constitutes a barrier to knowledge. The explanatory task is two-fold. Firstly, the task is to explain why it is the case that if vague expressions have sharp cut-off points, then we cannot know them. The second challenge is to explain why we lack knowledge in borderline cases, e.g. when Michael is a borderline case of baldness, then we cannot know whether he is bald or not.

The Williamsonian explanation of ignorance due to vagueness makes use of independently plausible epistemic principles to show that ignorance of cut-offs and ignorance in borderline cases are to be expected. In particular, it makes use of the safety condition on knowledge and the model of inexact knowledge to explain ignorance due to vagueness. It is no surprise that this part of the epistemicist story about vagueness has come under attack: if the explanation were good, it would make epistemicism very attractive.

There are a few epistemological aspects of the epistemicist model of vagueness that have come under attack including margin for error principles and the safety principle. In what follows I will outline some of the important critiques of the Safety Principle.

5.1 Safety and Margins for error

Williamson (1994b) makes use of the established safety-based principles in epistemology to explain ignorance due to vagueness. A merely true belief cannot constitute knowledge if the belief is not safe from error: safety from error is a necessary condition on knowledge. On a simple model, the safety condition is given by:

> (Simple Safety) X's belief that p is safe if and only if in all close possible worlds, it is not the case that X believes that p and p is not true.

Some theorists treat safety as a necessary and sufficient condition for knowledge; however, all that is required for the epistemicist theory to work is for safety to be a necessary condition for knowledge.

The safety condition for knowledge allows us to deal with the standard Gettier cases. For instance, in the *broken clock case*, X comes to have a belief that it's 7 o'clock because the clock is showing that it's 7 o'clock, but the clock is broken; in fact, it is 7 o'clock, so the X's belief is true and justified (looking at a clock is a standard way of gaining knowledge about what time it is). However, the belief is not safe: had the time been slightly different X would still believe that it's 7 o'clock, but the belief would have been false.

The safety condition allows us to come up with a model for inexact knowledge. Suppose that I am judging the height of a tree in the distance. My knowledge of the height of the tree is inexact. I can know that the tree in the distance is not 2 meters high and I know that it is not 200 meters high (its height is somewhere in between the two values). However, I don't know the exact height of the tree, i.e. for no N do I know that the tree is exactly N meters in height. The reason why I don't know the exact height of the tree is the fact that my perception is imperfect and requires margins for error of the form:

> (MFE-P) For all N, if X knows that the tree is at least N meters in height, then the tree is at least N+r meters in height (where r is the relevant margin for error).

In most of the discussion that follows, I will assume that the margin for error is 1 for convenience.

The epistemicist claim is that semantic plasticity produces failures of safety of a particular kind: failures of safety *due to* semantic plasticity. There are multiple factors that may produce the failure of safety. An analogy might be useful here. Suppose that you are still trying to determine the height of the tree, but now you have a yardstick to help. You determine that the height of the tree is exactly equal to ten times the length of your yardstick. In such a case, the initial source of inexactness of our knowledge, namely our perceptual limitations, has been eliminated. But suppose that the person who makes the yardsticks is a bit sloppy: some of the sticks he produces are shorter and some are longer. Your belief that the tree is 10 yards in height would not be safe in such a case, because had you gotten a longer stick, you would have concluded that the tree was 9 yards in height.

In case of a vague concept or word, the epistemicist claims that the reference of that concept or word shifts undetectably across close possible worlds. If it is true that Michael is bald, but he is only borderlinely bald, then my belief that Michael is bald is unsafe. Even if Michael counts as 'bald' relative to how it's actually used, it does not count as 'bald' relative to how it's used in a close counterfactual world. Shifts in the meanings of our concepts and words are conceptual sources of error. Since knowledge requires that we are safe from all kinds of errors, the safety condition for knowledge has to take these into account. In Chapter 4 I expand on the notion of safety relative to a factor.

Safety relative to semantic plasticity requires slightly different margins for error. Knowledge of meanings of our words is also inexact.[31] For example, if I

[31] That does not mean that our knowledge of meanings of linguistic expressions is lacking, as argued by Keefe (2000) and Tye (1997). As English speakers we understand the word 'bald' perfectly well (Williamson 1997b), e.g. there is nothing that we could learn from others (in contrast to a situation where one is not a speaker of the language). In particular, we know that 'baldness' denotes baldness, that 'that

order men that differ in height by 1 mm in a row, with the first man being clearly not tall and the last man being clearly tall, then for no man do I know that he is the first tall man in the row. However, the reason for my ignorance is not perceptual: we assume that I may measure the height of each man as precisely as possible; the reason is that my knowledge of the concept of 'tall' is inexact. Thus, if I know that a certain man is tall, then it must be the case that if the intension of the word 'tall' had been slightly different, then the man would still be in the extension of 'tall' in that circumstance. Suppose that 'tall' refers to those over N cm in height and 'tall*' to those over N+i cm in height. Then for all i that are less than the relevant margin for error, it is the case that if I know that X is tall, then X is tall*.

The Williamsonian margins for error have been heavily criticised in the literature. To be comprehensive in my presentation of the debate on epistemicism, I outline the main issues of contention below. However, in my opinion Williamson has been successful in answering his critics. Thus, I will not discuss these issues in later chapters.

5.2 Margins for error – Gómez Torrente (1997; 2002) & Graff Fara (2002)

5.2.1 Formal epistemicist model

The general fixed margin epistemicist model of margin for errors is a quadruple ⟨W, d, r, [.]⟩ such that (Williamson 1994b):

- W is the set of worlds in the model
- d is a metric on worlds such that $d(x, y)=d(y, x)$

John is bald' means that John is bald, and we know the penumbral connections that 'bald' features in (e.g. if a man with N+1 hairs is bald, then so is the man with N hairs).

- r is the non-negative margin for error in the model
- [.] is the standard interpretation function mapping the formulas in the language to subsets of W such that
  - $[\neg P] = W - [P]$
  - $[P\&Q] = [P] \cup [Q]$
  - $[KP] = \{w \in W: \forall x \in W\ (d(x, w) \leq r \rightarrow x \in P)\}$.

The interpretation of the model is straightforward: it can be interpreted as a standard modal logic model with K taking the place of $\Box$. W is the set of possible worlds differing with respect to the cut-off points that are fixed to vague expressions. d can be interpreted as a similarity metric on worlds. Where R is any symmetric accessibility relation between the worlds in the model, d can be defined as:

- $d(x, x)=0$
- $d(x, y)=1$ if xRy but $x \neq y$
- $d(x, y)=2$ if not xRy

The margin for error r is assumed to be fixed. The operator K can be interpreted as saying 'it's not unknowable for the reasons that have to do with vagueness that' and as required by the safety conception the sentence 'KP' is true at a world w only if P is true in all worlds that are within the margin of error from w. Since d is symmetric and r is fixed throughout the model, the model validates the B-axiom of modal logic: $\neg P \rightarrow K \neg KP$. However, there are also variable margin models where B-axiom fails.[32] Such a model is also a quadruple (W, d, r, [.]) and satisfies the same clauses as the fixed margin model apart from the clause for K, which in the variable margin models is given by:

$[KP]=\{w \in W: \exists \delta > r\ \forall x \in W\ (d(x, w) \leq \delta \rightarrow x \in P)\}$.

[32] See the Appendix of Williamson (1994b) for a proof of this result.

Let's consider challenging arguments to the margin for error model due to Gómez-Torrente (1997;2002) and Graff Fara (2002).

5.2.2 Transparent propositions lead to known cut-off points

Mario Gómez-Torrente (1997) argues that margin for error models have an interesting anti-epistemicist consequence under some plausible assumptions. Let M be a fixed margin model such that:

- $W=\{s_n: n \in \mathbb{N}\}$
- $r=1$
- $s_i \vDash B(n)$ iff $n<i$
- $s_i \vDash KP$ iff $\forall s_j \in W\ (d(s_i, s_j) \leq 1 \rightarrow s_j \vDash P)$

A sentence 'B(n)' is to be interpreted as 'a person with n hairs is bald' and 'KP' is to be interpreted as 'it is knowable that P'. Furthermore, *m* reiterations of K will be shortened to 'K^m'. Thus, for any *m* the margin for error principle states:

(1^m) $\forall n(K^{m+1}B(n) \rightarrow K^mB(n+1))$

Suppose that we are evaluating sentences relative to world s_i. Recall that $s_i \vDash B(n)$ if and only if $n<i$. Then by the margin for error principle $s_i \vDash K^mB(0)$ if and only if $m<i$. For suppose that $m \geq i$. Then by repeated applications of (1^m), we get $s_i \vDash K^{m-i+1}B(i-1)$. But this is false, because $d(s_i, s_{i-1}) \leq 1$ and $s_i \nvDash B(i)$. So for any $n>0$, $s_i \vDash K^nB(i-1)$ is false. Thus $s_i \vDash K^mB(0)$ only if $m<i$.

Gómez-Torrente (1997) argues that this result is problematic for epistemicism, because it disallows indefinite reiterations of K. However, sentences such as 'B(0)' are transparent: it's analytic (or very close to analytic) that the man with 0 hairs is bald. There does not seem to be any way in which we could be wrong about that. So it should be possible to reiterate as many Ks in front of 'B(0)' as we like; but the epistemicist model disallows that. If the cut-off point for baldness is n, then only n-1 reiterations of K are allowed.

Williamson (1997b) argues that limit on possible reiterations of K is plausible. Even if it is analytic that a man with 0 hairs is bald it is not analytic that it is known that the man with 0 hairs is bald. Indefinite reiterations of K are only plausible if we think that the KK-principle holds (for any p, if Kp then KKp) or if we at least accept intuitions that lead to the adoption of KK (even if we reject KK in full generality), i.e. the fiction that our mental states are transparent to us and that there is no loss of transparency with every reiteration of K.

Therefore, the Williamsonian epistemicist seems to have a way out by denying that any vague sentences are transparent in the way that Gómez-Torrente requires them to be for his argument to go through. But Williamson (1997b) argues that epistemicists can be more ambitious and accommodate the intuition that for some vague sentences such as 'B(0)', K can be reiterated indefinitely. The way that this is done is by introducing cropped models where the initial sequence of worlds is removed from the model. For example, if we wish 'B(0)' to be transparent, then 0 is cropped from the model so that $W=\{s_n: n\in\mathbb{N} \ \& \ n>0\}$; all remains the same apart from the fact that indefinitely many reiterations of K for 'B(0)' are allowed in the model and that the margin for error principle is reworked to state:

(ME) If $s_i\vDash KP$ and $d(s_i, s_j)\leq 1$, then $s_j\vDash P$.

However, the problem with the cropped models is that for any world in the model there is a knowable cut-off point (Gómez-Torrente 2002). In particular it can be shown that for any s_i there is some m such that $s_i\vDash K(K^mB(0)\&\neg K^mB(1))$. There are a few ways to show this result. The simplest is to use the B-axiom: $\neg P\rightarrow K\neg KP$, which holds in all fixed margin models (Graff Fara 2002). Suppose that only s_0 is cropped from the model. Then for some large $m>i-1$, by the margin for error principle, $s_i\vDash\neg K^mB(1)$.

By B-axiom, we get $s_i \vDash K\neg K^m B(1)$. Since s_0 is cropped from the model $s_i \vDash K^n B(0)$ holds for any n. In particular it holds for n=m+1, so $s_i \vDash K^{m+1} B(0)$ holds. Given that KP&KQ implies K(P&Q), we get $s_i \vDash K(K^m B(0) \& \neg K^m B(1))$.

Interestingly, moving to variable margin models does not get us out of trouble. Though the B-axiom does not hold in variable margin models, for all models that are stepping-stone models, i.e. models for which the following holds:

> for any w and v if (d(w, v)=n$\cdot r$ for some n≥2), then $\exists u_1 \exists u_2 \exists u_3 \ldots$($d$(w, u_1)≤r & d(u_1, u_2)≤r & … d(u_3, v)≤r),

the following weaker principle holds (Graff Fara 2002):

> B*=¬P→K¬KKP

Clearly, stepping-stone models are precisely models that are of interest to us when discussing the Sorites Paradox. However, it can be shown (by an argument analogous to the one presented above) that B* also lead to the same result: for any s_i for some sufficiently large m it is the case that $s_i \vDash K(K^m B(0) \& \neg K^m B(1))$.

Can the epistemicist retreat back to claiming that in fact for no n the sentence 'B(n)' is transparent? This would allow us to go back to the uncropped models where such worrying results do not seem to arise. However, as shown by Williamson (2002a), very similar results can be also shown to hold in uncropped models as well. Let M be our uncropped model again. Recall that $s_i \vDash B(n)$ holds if and only if n<i. Then $s_i \vDash B(0)$ holds for all i>0. But note that $s_i \vDash \neg K\neg B(n)$ holds for all n<i+1. So $\neg K\neg B(0)$ holds in all worlds in the model. So for any m, it is the case that $s_i \vDash K^m \neg K\neg B(0)$. Furthermore, note that for all n>0 it is the case that $s_i \nvDash \neg K\neg B(n)$ if and only if n>i. Consequently, $s_i \nvDash K^m \neg K\neg B(n)$ if and only if n+m>i. So for a large m such that m>i it is the

case that $s_i \vDash K\neg K^m\neg K\neg B(1)$, because m+1>i and m+1>i−1 and m+1>i+1. Therefore, for any s_i there is an m such that we get $s_i \vDash K(K^m\neg K\neg B(0)$ & $\neg K^m\neg K\neg B(1))$. The result generalises: for any s_i and any x there are m, l such that $s_i \vDash K(K^m\neg K^l\neg B(x)$ & $\neg K^m\neg K^l\neg B(x+1))$. These results definitely reveal some interesting structural features of the epistemicist models used by Gómez-Torrente, Graff Fara, and Williamson.

However, such problems arise as a result of adopting a simplified model for analysis. If we try to raise these problems in a more complex model, they do not seem to arise. Suppose that instead of having points of evaluation corresponding to natural numbers, we use the standard possible worlds framework. Furthermore, suppose that the semantics for 'B' just vary with the way that the semantics for 'bald' vary across the space of worlds. There are worlds where the meaning of 'bald' is very far from our own, e.g. where 'bald' means red. On such a model there will be no sentences like '$\neg K\neg B(0)$' above that would be true in every point in the model. Once the model is more complex (and realistic) such artificial problems do not arise.

5.3 Range estimates are a better model for inexact knowledge (Mott 1998).

Williamson uses the intuitiveness of perceptual margins for error (that apply in cases such as when I estimate the height of a tree) to argue for the conceptual margins for error (that apply in borderline cases) which are crucial for his account of vagueness. Peter Mott (1998) presents an argument against the perceptual margin for error principles. The force of the argument is that even though the epistemicist solution requires the conceptual margins and not the perceptual ones, Williamson persuades us to accept the conceptual margins by analogy with the perceptual ones which are simpler and more intuitive. So

showing that MFE-Ps are incorrect would damage the epistemicist case by relieving it of its persuasive force stemming from the claim of ideological conservativeness.

Mott's argument is that our judgements in cases of inexact knowledge are not like Williamson says they are. When we judge a height of a tree from a distance, using our estimate of its height and the margins for error for our judgements, we can come to know that the height of a tree is within some range, e.g. we say 'the tree is between 10 and 20 meters in height'. Suppose that my margin for error in judging the height of the tree is 5 meters. I estimate that the tree is 15 meters in height; since I know my margin for error, I know that the tree is between 10 and 20 meters in height. Suppose that the tree is actually 20 meters in height. Since the height of the tree is within the range <10, 20>, my judgement was just right, and I can still have knowledge that the tree is between 10 and 20 meters in height (even though my estimate was slightly off). But on Williamson's model I don't know that the tree is between 10 and 20 meters in height. For instance, on Williamson's model, I am not in a position to know that the tree is not 21 meters in height (for a margin of error at least 1 m), if it is in fact 20 meters in height. However, Mott claims, if the tree is 20 meters in height, I can know that it is between 10 and 20 meters in height.

Williamson (2000b) replies to Mott's argument by noting that our knowledge of our margins for error is inexact and thus also associated with margins for error.[33] Suppose that I can inexactly know my margin for error in

[33] Mott (1998) presents a second argument which is supposed to show that Williamson's margin for errors lead to a sceptical result. In particular, he aims to show that Williamson's margin for error principles are a serious obstacle to knowledge of other people minds. That argument suffers from the same flaw as the argument outlined above: it presupposes that we can know what the margins for error are with perfect accuracy; once that presupposition is removed, the argument no longer works.

judging a tree with a margin for error of 1 meter. Then if I know that the tree is between 10 and 20 meters, then the tree is between 11 and 19 meters. So Mott's case does not hold up: if the tree is 20 meters in height, we are not in a position to know that it is between 10 and 20 meters in height. The objection could work under the assumption that we are in a position to know exactly what our margins for error are. This is unargued for (not to mention independently unreasonable), so Mott's objection should be rejected.

5.4 Complex predicates invalidate margins for error (Sorensen 2007).

Roy Sorensen (2007) argues that Williamson's margin for error principle work for simple predicates, but fail for complex predicates, which makes the Williamsonian story about the origin of the ignorance due to vagueness implausible. Sorensen attributes to Williamson margin for error principles that are of relevance to vagueness of the form:

> (MFE-Sorensen) Where 'B' is a vague predicate, if it is known that B(N), then B(N+1).

Sorensen claims that this principle works for simple predicates such as 'bald'. But it fails for complex predicates such as 'small or equal to 10'. Suppose that it is borderline whether 10 is small and that 10 is the last small number. Nevertheless, it is known that 10 is either small or equal to 10, because it is known that 10 is equal to 10. So:

> (K10) It is known that 10 is small or equal to 10.

Clearly the predicate 'is small or equal to 10 is vague', because it has borderline cases due to the vagueness of the first disjunct. So by MFE-Sorensen we get:

(K11) 11 is either small or equal to 10.

But 11 is neither small (10 is the last small number) nor equal to 10. Therefore, the margin for error principles that Sorensen attributes to Williamson fail.

Williamson responds by noting that MFE-Sorensen principles are simplified versions of his own margin for error principles. In the general form, the relevant margin for error principles take the form:

(MFE-W) Where 'B' is a vague predicate, if it is known that B(N), then B*(N).

'B*' denotes the intension of the predicate 'B' in close possible worlds, where the meaning of 'B' is slightly different. MFE-W reduces to MFE-Sorensen if it is the case that:

(Reduction) If B*(N), then B(N+1).

This holds for simple predicates, but not for more complex ones. Suppose that 'small*' denotes the intension of 'small' in a close world and that the following holds:

(Synonymy) For all x, x is small* if and only if x+1 is small.

MFE-W applied to the predicate 'is small or equal to 10', yield:

If it is known that N is small or equal to 10, then N is small* or equal to 10.

By Synonymy we get:

If it is known that N is small or equal to 10, then N+1 is small or N is equal to 10.

But this does not entail K11, so Sorensen's argument does not work.

6 Semantic Issues

6.1 Defining Definiteness

The epistemicist account of borderline cases has recently been shown to need further explication.[34] There are two general problems with the account. Firstly, there is a problem with the way that the definiteness operator Δ is defined: it seems that the definition is inconsistent with the safety explanation for ignorance due to vagueness. Secondly, there is a problem with the interaction of the epistemicist definiteness operator with modal operators. The problems are quite technical, so I will outline them only briefly in this section; I will attempt to explicate them in greater detail and answer them in Chapters 4 and 5.

Michael Caie (2012) presents two ways of interpreting the epistemicist definiteness operator Δ, which are given by (O), the other-worldly account, and (A), the actualistic account, below:

> (O): ⌜ $\Delta\phi$ ⌝ is true at @ if and only if in all possible worlds w that are close to @, ϕ as used at w is true at w.
> (A): ⌜ $\Delta\phi$ ⌝ is true at @ if and only if in all possible worlds w that are close to @, ϕ as used at w is true at @.

It seems that (A) is more in the spirit of the epistemicist story. Ignorance due to vagueness is supposed to have a specific source in shifts in the meaning of concepts and words. However, (O) seems to allow for factors other than semantic plasticity to produce ignorance. Suppose that Michael is definitely tall, i.e. his height is further from the cut-off point for 'tall' than required by the relevant margin for error. But suppose also that in a nearby world w^* the

[34] See Caie (2012), Magidor (2018), Yli-Vakkuri (2016) for objections of this kind.

standards for being 'tall' are slightly stricter (fewer people count as tall) and that in such a world Michael is slightly shorter than he actually is. The compound of the change in Michael's height and the change in the meaning of 'tall' is greater than the relevant margin for error and Michael is not in the extension of 'tall' at w^* as it is used at w^* (even though the shift in meaning alone is within the margin for error so it counts as relevantly close). This is a problem, because Michael's height was supposed to be within the relevant margin for error. On views such as (O), these cases are problematic because (O) does not take into account that close worlds may differ slightly on a variety of dimensions; however, in discussing vagueness we are only interested in one particular dimension of shifts across possible worlds.

Adopting (A) is a way of picking out the factor that is relevant to vagueness. We are holding fixed the precise facts in the actual world and allowing only for variation along the dimension that is relevant to vagueness: shifts in reference due to semantic plasticity. However, such accounts fall into trouble as well. For one thing, there are inconsistent with the spirit of the safety explanation for ignorance. The safety condition on knowledge requires that we would be safe from error in close possible worlds, so we should only count a sentence as definitely true if it is not the case that if we were in a close world, we would make a mistake; but this is not required by (A).

Furthermore, it is difficult to provide an account of the definiteness operator that would interact well with the modal operators. For example, (O) invalidates the following intuitive principle: necessarily, if ϕ is definitely true then ϕ is true $\Box(\Delta\phi \rightarrow \phi)$. On the other hand, some versions of (A) still invalidate $\Box(\Delta\phi \rightarrow \phi)$ and others violate equally intuitive principle: it's definitely true that ϕ is actually true if and only if ϕ is true $\Delta(A\phi \leftrightarrow \phi)$, where 'A' stands for the actuality operator). I will outline the details of the problem in Chapter 5. For now, what we know is that there is more work to do for the

epistemicist; we have to spell out the details of the account of borderline cases, including the semantics of the definiteness operator, that would avoid the subtle problems that have been developed in the literature.

7 Puzzle cases

7.1 Speech Reports Puzzle – Dorr & Hawthorne (2014)[35]

Use of language varies between speakers. It is rarely the case that the use of language of the speaker and the hearer are identical. If so, how is communication possible? Williamson's (1997; 1999) standard response is to say that in cases of successful communication we can assume that the speaker and hearer belong to the same linguistic community. The linguistic meaning of utterances of the speaker and the hearer are parasitic on the use of 'bald' in the linguistic community. If this is the case, then communication failure does not take place.

However, the problem does not go away in case of counterfactual speech reports as demonstrated by Dorr and Hawthorne (2014). Dorr and Hawthorne show that casual counterfactual speech reports such as 'if Sally had come to dinner with us, she would have said that salad is delicious' are false since the proposition referred to by the that-clause in the speech report is distinct from the proposition that would have been expressed by Sally due to the semantic plasticity of the word 'salad'. But it seems very unlikely that we would be systematically making a mistake in making such utterances.

[35] To the best of my knowledge, the observation that epistemicism may lead to problems with validating intuitively true examples of speech reports was made first by McGee & McLaughlin (2004).

Note that this is not only a problem for epistemicism, but for anyone that thinks that our language is semantically plastic. I do not address the puzzle in the thesis; it will have to be resolved in later research.

7.2 Omniscient Speaker Puzzle – Sennet (2012)

Suppose that Barney, the Omniscient Linguist, is a member of some linguistic community. In Barney's community the boundary for counting as 'rich' is $40,000. Barney has cooked up a scheme to stabilise the intension of 'rich' in his community. Whenever someone uses the term 'rich' in a way that would change the intension, Barney makes sure (e.g. by making a second utterance that would cancel out the effect of the first utterance) that the change does not occur. It seems that the boundary of 'rich' is safe: it does not shift across nearby worlds. However, Sennet (2012) argues, 'rich' is still vague in the community. If according to epistemicism vagueness is related to the propensity of intensions of expressions to shift across nearby worlds, this is a problem for epistemicism. I deal with this puzzle in Chapter 7.

7.3 Persons Puzzle – Hawthorne (2006)

Suppose that there is a speaker of a non-actual, but close variant of the language (call the variant 'Twinglish') who is sitting in a chair. Call the speaker 'Arthur'. There is a speck of dirt on his face. In English, the actual variant of the language, 'person' refers to Dusty: a fusion of the speaker and the speck. In Twinglish, 'person' refers to Clean: the speaker without the speck. Suppose that Arthur utters the sentence 'I am a person'. Hawthorne (2006) shows that very intuitive principles lead us to the conclusion that 'I am a person' is borderline. In particular, 'I am a person' seems to be false as uttered by the Twinglish speaker. There is one and only one speaker in the chair. That speaker

is a person. Since 'person' in English refers to Dusty, Dusty is the speaker. Since 'I' is used to refer to the speaker, 'I' refers to Dusty. But 'person' in Twinglish refers to Clean, who is distinct from Dusty. So the Twinglish utterance is false. Since Twinglish is a nearby variant of the language, this poses a problem for epistemicism, as it would make 'I am a person' borderline on the epistemicist account. I don't address this puzzle in the thesis.

8 Conclusion

My aim for the chapter was to state the theoretical commitments of epistemicism and outline the key points of disagreement in the debate about epistemicism. I intend to answer the challenges and develop the epistemicist case in following chapters.

Part I – Metasemantics

The first part of the thesis deals with the metasemantic issues that epistemicism faces. Epistemicist metasemantics has been the main focus of the debate over epistemicism following the publication of *Vagueness* in 1994. Chapter 2 outlines the general challenges to epistemicist metasemantics as well as responses to critics from Timothy Williamson. Chapter 3 deals with a more recent challenge to epistemicist metasemantics due to Miriam Schoenfield (2016). Schoenfield argues that epistemicist treatment of moral vagueness is incompatible with moral realism. My aim in the chapter is to show that Schoenfield's argument against epistemicism rests on a misreading of the theory. I also present a version of epistemicist metasemantics that is not vulnerable to the problems that Schoenfield points to.

Chapter 2

Metasemantic Problems for Epistemicism

ABSTRACT

The majority of space and attention in the debate on epistemicism immediately following the publication of *Vagueness* in 1994 by Timothy Williamson has been devoted to the discussion of epistemicist metasemantics. The general sentiment expressed in this literature is that epistemicism is an elegant theory, which nevertheless is untenable because of its crazy metasemantic commitments. The aim of this chapter is to provide an overview of the debate over epistemicist metasemantics and to present some of the lines of response that have been given by Timothy Williamson in various replies to the critics. I also argue that there remain interesting metasemantic questions for the epistemicist to answer.

WORD COUNT: 8,630

1 Introduction

In the debate about epistemicism that emerged after the publication of *Vagueness* by Timothy Williamson in 1994, several authors expressed doubts

about the plausibility of the epistemicist metasemantic story. The doubts about the epistemicist account concern multiple dimensions of the theory, but they can be summarised by the following claim:

> (Metasemantic Challenge) The epistemicist account of how vague expressions get assigned meanings is implausible.

Usually, the argument goes as follows. Epistemicism has many nice features as a theory, but the benefits are not worth the costs of implausible metasemantic theory. Thus, we should reject epistemicism. My hunch is that this sentiment is expressed by the famous incredulous stare that often faces epistemicism. How on earth could it be the case that our non-technical and sloppily used words have sharp cut-off points?

The epistemicist has several options to counter the challenge. The 'minimum effort' option is simply to say that the advantages of epistemicism simply outweigh the costs of leaving certain questions in metasemantics open. Epistemicism is a good theory as far as the things it says something about go (logic, epistemology); it may very well be incomplete, but Rome was not built in a day. After all, theories alternative to epistemicism also rely on metasemantic generalisations. For instance, the supervaluationist has to rely on the metasemantic generalisation that use of any vague expression does not determine a single intension, but that there are multiple intensions compatible with the use facts. Thus, reliance on metasemantic generalisations (in a situation where we don't have a working metasemantic theory) is not unique to epistemicism. If epistemicism is on a par with other theories as far as reliance on unsupported postulates in metasemantics goes, then epistemicism should be the theory we go for given the advantages that epistemicism over the alternatives (classical logic, dealing with higher order vagueness). There are

merits to this answer though arguably it has little power to persuade the critics of epistemicism. Therefore, the critiques have to be dealt with more directly.

My aim in this chapter is to outline the main lines of criticism of epistemicist metasemantics alongside some defences to these objections (mounted primarily by Williamson). The goal is not to settle all these issues here: the breadth of the literature does not allow me to do justice to all these arguments. I will have to settle for something more modest. What I intend to show is that the epistemicist can offer an initial response to the problems raised in the literature. In the next chapter, I will discuss in greater detail and offer a solution to what I think is the most difficult problem for epistemicist metasemantics coming from moral vagueness.

In Section 2, I will outline the metasemantic commitments of epistemicism. In subsequent sections I show how different aspects of the epistemicist metasemantics have been put to question by the critics. I also outline briefly the lines of response that have been provided by Timothy Williamson that in my opinion give the epistemicist the best chance of responding to the critics. I will conclude with a section on questions that, in my opinion, present the biggest challenge to epistemicist metasemantics, which I will aim to answer in Chapter 3.

2 Metasemantic Commitments of Epistemicism

Epistemicism identifies semantic plasticity as a characteristic feature of vague expressions. Using the safety condition on knowledge, the epistemicist then shows that semantic plasticity produces ignorance in borderline cases. Therefore, the claim that vague expressions are semantically plastic is a core tenet of Williamsonian epistemicism: the ultimate explanation of ignorance in

borderline cases relies on it. The Master Argument for semantic plasticity of vague expressions can be summarised as follows:

(1) Vague expressions have sharp boundaries.

(2) If vague expressions have sharp boundaries, then they are semantically plastic.

(3) Vague expressions are semantically plastic.

The argument is valid.[36] The attacks on epistemicism come in many forms. Some focus on arguing directly against (1) or (2). Some critics focus on undermining the arguments provided by Williamson and others in favour of (1) or (2); consequently, they argue that Williamson has not succeeded in establishing (3), as there are no reasons to believe the premises once the arguments for them have been undermined.

Let us first consider the arguments for (1) and (2). Then I will present the issues pointed out by critics and I will show ways that these issues can be resolved.

2.1 Arguments for (1)

Epistemicism is committed to classical logic (including meta-rules such as reduction or conditional proof) and the principle of bivalence:

> (Bivalence) If u says that p, then u is either true (in case that p) or false (in case that $\neg p$).

[36] Some critics argue directly against (3): they claim that the epistemicist has not provided independent reasons to think that meanings are semantically plastic (Keefe 2000, Horwich 1997). This is problematic, because all the epistemicist has to show is that (1) and (2) are true to establish (3). If the critic wants to reject the epistemicist case for (3) they have to argue against (1) or (2) (or the epistemicist case for (1) and (2)).

Bivalence guarantees that vague predicates have sharp cut-off points. By the least number principle, there is a smallest amount of money that one could have to count as rich, there is a unique smallest number that counts as large etc.

There are two arguments in favour of sharp boundaries that Williamson provides.

Argument from logic. Humans are better at logic than philosophy (Williamson 1997a). We should go with our best logical theories if there are no knockdown reasons to abandon them. Since classical logic, which for Williamson includes bivalence and the classical meta-rules, is most elegant and powerful (and used widely by mathematicians) we should retain it as a default option.

Argument from truth. Very fundamental principles that the notion of truth satisfies can be shown, together with classical logic, to entail bivalence. Since bivalence entails sharp boundaries, the argument establishes that there are sharp boundaries. Truth is one of the most fundamental and basic philosophical notions. Lessons from the study of semantic paradoxes like the Liar indicate that truth gives rise to very complex problems. Despite these complications, there are some things that we know about truth. One of the things we know about truth is that it applies to utterances that say that something is the case if and only if what the utterance says is the case, is the case. The thought can be summarised as follows (Williamson 1994b):

> (Truth) If an utterance u says that p, then u is true if and only if p.
> Similarly, the notion of falsity is guided by the following principle:
> (Falsity) If an utterance u says that p, then u is false if and only if not p.

The above principles are necessary conditions for truth and falsity. They seem to obvious that they come very close to being truisms. If we can say anything about the notions of truth and falsity, and it seems that we should say something if the notions are to be intelligible at all, we can say that Truth and Falsity characterise these notions. Williamson (1994b) shows using classical principles (and the classical meta-rule of argument by cases) that Truth and Falsity entail bivalence. Suppose that u says that p. Classical logic tells us that law of excluded middle holds, so either p or not p. Suppose that p. Then, by Truth, u is true. By the rule of disjunction introduction, u is either true or false. Suppose on the other hand that not p. Then, by Falsity, u is false. By disjunction introduction, u is either true or false. Discharging the initial hypothesis, we get Bivalence: if u says that p, then u is either true or false.

2.2 Arguments for (2)

There are three arguments for (2) that we can identify in the literature.

Argument from Plurality. This is an ingenious argument developed by Dorr & Hawthorne (2014). The authors use the phase space model and measure theory, to establish semantic plasticity. The basic idea is that there is simply so many possible intensions that a vague expression could be assigned, that there is no region in the logical space where the intension is stable. I will not go through the details of the argument since it's relatively new and not yet widely discussed in the literature. Nevertheless, it's an argument that opponents of epistemicist metasemantics have to contend with.

Argument from Parity.[37] Suppose that the actual boundary for counting as 'tall' is 180 cm. There is a possible world w^* (perhaps not a close one) where

[37] Here's Williamson's (1997b, p. 948-949) argument for a continuous change in reference across the logical space: "And what is the alternative to the postulated correspondence between shifts in use and shifts in reference? The reference of a vague

the boundary is much higher, e.g. 190 cm. However, then there are also worlds where the boundary is set at some value between 180 and 190 cm. Then holding everything else constant, such worlds will be more similar to the actual world than w^* is. Therefore, there is a series of worlds between the actual world and w^* where the boundary gets higher and higher as we progress through the series. It is possible that the jump of the boundary from 180 cm to 190 cm would not occur continuously in the way just described. However, then it seems that 180 and 190 cm would be natural boundaries of sorts. This is implausible since the assumption made in borderline cases is that none of the candidate referents is more natural than others.[38]

Argument from Use Metasemantics. The third argument is based on the toy metasemantic model presented by Williamson (1994b). Presumably, the meanings of the words in our language have something to do with the way we as a linguistic community use them. Therefore, let's suppose that usage

---

term would be insensitive to most small shifts in use, but would change in large jerks as the pattern of use shifted across critical boundaries. The privileged extensions would mark out something like hidden natural kinds. Situations like the following would obtain. 'Bald' actually refers to the property of having fewer than 3,832 hairs on one's scalp. If we had been as likely to apply 'bald' to someone with N hairs as we actually are to apply it to someone with N + 100 hairs (for all N), then 'bald' would still have referred to the property of having fewer than 3,832 hairs on one's scalp. But if we had been as likely to apply 'bald' to someone with N hairs as we actually are to apply it to someone with N + 1,000 hairs (for all N), then 'bald' would have referred to the property of having fewer than 2,832 hairs on one's scalp. 3,832 and 2,832 are natural boundaries of a sort; 3,732 is not. That is implausible, although not incoherent."

[38] This point is sometimes misunderstood in the literature. For instance, Horwich (1997) argues that Williamson (1994b) does not present indubitable arguments for (2): it is possible that meanings are not shifty as epistemicism predicts. Williamson (1997c) correctly responds to Horwich that this is not a reasonable standard by which to measure epistemicism. We should not expect epistemicism to establish (2) as indubitable. Instead we should simply apply normal scientific standards: the epistemicist claim that meanings are semantically plastic given that they are sharply bounded is just more likely than the alternative.

determines the meanings of vague expressions in our language. The usage of expressions within a linguistic community varies. It could easily be the case that our dispositions to use language would be different; it could easily be the case that some utterances that were actually made would not be made (and analogously some that were not actually made could easily be made). For instance, our use of the word 'tall' could easily be slightly more conservative so that we would count less people as tall. It is natural to suppose that had our usage been slightly more conservative, the boundary for 'tall' would also be slightly higher; therefore, it's natural to suppose that small shifts in usage produce small shifts in the intension of vague expressions in our language.

Note that unlike the previous argument, this argument makes explicit claims about the metasemantic mechanism which determines meaning. The point of the Argument from Parity was that even if we treat the metasemantics a 'black box', i.e. something about the workings of which (input factors, ways in which these inputs determine the output) we know nothing about, it is still more likely that if expressions in our language have sharp boundaries, then these boundaries are plastic. On the other hand, the claim of this argument is that the epistemicist can produce a compelling toy metasemantic model, on which meanings supervene on use, which would explain how meanings are shifty. This is important because even if we find fault with the usage-based toy metasemantic model, the epistemicist can fall back on the Arguments from Parity and Plurality.

In the next section, I will outline the objections that critics of epistemicism had with the above commitments.

3 Objections to (1) - sharpness of vague predicates

The first line of attack that can be found in the literature is aimed at the epistemicist claim that vague expressions have sharp boundaries.

3.1 Direct objections against (1)

3.1.1 We don't use vague expressions as if they are sharply bounded

The first line of attack against the epistemicist focuses on the way we use vague expressions. Epistemicism claims that vague expressions have sharp boundaries. But we don't seem to use them in this way.

Firstly, it seems that in case of vague predicates such as 'bald', there are multiple intensions that are consistent with our use (Keefe 2000, Chapter 3). In most series, there seems to be a class of precise points that could serve as a cut-off; these are the borderline cases. Therefore, it seems that our use of vague expressions selects multiple candidate intensions as equally eligible and not just one. Thus, demanding that vague expressions have sharp cut-off points is to sever the connection between meaning and use.

Secondly, demanding that expressions have sharp cut-off points goes against the grain of how casually and sloppily we use vague expressions (Machina & Deutsch 2002; McGee & McLaughlin 1998). It would be miraculous if our casual use of words such as 'bald' managed to attach to a single intension.

The following analogy may be useful in thinking about the epistemicist case for sharp boundaries. Suppose that we construct a standard roulette machine with slots from 0 to 36.[39] The slots on the machine are constructed in such a

[39] If you prefer the American version of the roulette machine feel free to imagine that the roulette also contains a 00 slot in addition to the 0-36 slots.

way that on every turn when the ball is rolled it always lands in one and only one of the slots. It is fairly easy to explain why on every turn the ball lands in one and only one of the slots: the ball and the machine are just constructed in a way that guarantees that the ball will not stop on more than one slot or somewhere between the slots etc. However, the explanation of why the ball landed in the slot that it did on a particular turn is much harder. To give a full explanation we would need to account for all the forces that act on the ball including friction, air resistance etc. Showing this would require substantive knowledge of physics as well as the possession of extremely precise measuring equipment. Therefore, there is a contrast between the explanation of why the ball always lands on one and only one slot that appeals to *global* features of the setup and an explanation of how the ball landed in a given slot on a particular turn, which requires getting into the details of the *local* features.

We can make a similar point about epistemicist metasemantics. The explanation for why vague expressions have sharp boundaries is pretty straightforward: this is what the Arguments from Logic and Truth achieve. This is established by appealing to global principles, e.g. ones guiding the notion of truth (Williamson 2002). This is quite different from explaining how the local metasemantic features of a particular expression fix the boundary in a given case. Sloppy usage of language can fix a sharp boundary just like sloppy push of the roulette ball will end with the ball landing in one and only one slot. Therefore, it seems that the epistemicist can withstand the initial attack by appealing to global features of the model to explain sharp boundaries.

3.1.2 Sharp boundaries are not real

Some authors (Keefe 2000, McGee & McLaughlin 1998) express scepticism about the metaphysical weight that the epistemicist places on the existence of sharp cut-off points. There are good arguments to accept the thesis that vague

predicates have sharp cut-off points: Williamson's argument shows that classical logic and intuitive disquotational principles together entail bivalence which requires that vague predicates have sharp cut-off points. Williamson's conclusion is that in order not to sever the link between use and meaning we should conclude that sharp cut-off points supervene on use. But, the objection goes, that would be to treat the epistemicist model too seriously: sharp cut-off points are merely lightweight conceptual truths that don't put any constraints on reality. So it is not the case that vague predicates are really sharply bounded: it is merely the case that there is a predicate (namely the truth predicate) that can be consistently added to the language that is bivalent; but that is not enough to guarantee that there indeed are sharp cut-off points.

Williamson (1994b) famously argues that supervaluationist theory has to dispose of the Tarskian schema: 'S' is true iff S. It is easy to see why: if truth is equated with truth on every admissible precisification (i.e. supertruth), then borderline sentences which are true on some admissible precisifications and false on others will invalidate the right-to-left direction of the schema. However, McGee and McLaughlin argue that this does not mean that the supervaluationist has to dispose of the disquotational theory of truth: they may add a disquotational truth predicate to the language alongside a determinacy operator to get the best of both worlds. On such a view, vague predicates have sharp cut-off points due to the fact that truth is disquotational, e.g. there is a sharp cut-of between the bald and the non-bald people, but no particular boundary is determinately the cut-off point. Thus, it is not the case that our use determines the sharp cut-off points for vague predicates.

The argument relies on a false claim that the existence of the cut-off point is a 'conceptual' or an 'analytic' truth that puts no constraint on reality. All truths are true in virtue of the fact that they mean what they mean and in virtue of the fact that what they say is the case. For example, the sentence

'Napoleon died on St Helena' is true in virtue of the fact that it means that Napoleon died on St Helena and the fact that Napoleon died on St Helena (Williamson 2007b). Similarly the sentence 'All ravens are ravens' is true in virtue of the fact that it means that all ravens are ravens and the fact that all ravens are ravens.

Suppose that the cut-off point for 'bald' is N hairs. The claim that 'A person with N hairs is bald and a person with N+1 hairs is not bald' is true in virtue of the fact that 'bald' denotes baldness and in virtue of the fact that a person with N hairs is bald and a person with N+1 is not. After all, if truth obeys the Tarskian schema the sentence 'a person with N hairs is bald' is true if and only if a person with N hairs is bald, so it is made true by a non-linguistic fact that a person with N hairs is bald. So it is not the case that for borderline cases of baldness nothing makes it the case that they are in the extension of 'bald': the fact that they have sufficiently small amount of hair is what makes it the case that they are in the extension of bald. Therefore, the deflationary routes to the claim that vague predicates have sharp cut-off points are unsuccessful.

I believe that the above reply is the quickest way to argue the point. However, there is a stronger version of the objection from supervaluationists that does not appeal to analyticity.

The thought is that Williamson, using his arguments from logic and truth, shows that bivalence holds. However, then he uses his metasemantic model to argue that usage determines the sharp boundaries. The latter move can be doubted by the supervaluationist. The basic thought is that the step from bivalence to sharp boundaries is not as obvious. The thought is that there are multiple metasemantic models that can account for bivalence. The model that the Williamsonian epistemicist chooses is one on which a pattern of usage

within a linguistic community admits only one sharp boundary for any vague expression. But this is not the only possible model that accounts for bivalence.

There are of course other metasemantic models that account for bivalence. For instance, on a view proposed by Breckenridge and Magidor (2012), there are multiple sharp boundaries that are consistent with usage within a linguistic community and the boundary is simply selected arbitrarily among those. However, this is not an argument against sharp boundaries, but rather an argument against the epistemicist Use Metasemantics. The epistemicist point is not that it is inconceivable to have a different metasemantic model, but just that the Use Metasemantics model is the simplest. After all, we agree that usage is something that determines the candidate sharp boundaries for any vague expression. Since we already accept usage as the meaning determining factor and we have established bivalence, the parsimony of explanation suggests that we should go for a model on which usage determines the sharp boundaries. So there are plausible reasons to accept the epistemicist explanation of sharp boundaries.

3.2 Objections to the Argument from Truth

This class of objections aims to show that Williamson's (1994b) Argument from Truth is faulty. If they were successful, then Williamson's claim for sharp boundaries would be weakened.

3.2.1 Williamson presupposes the disquotational theory of truth (McGee & McLaughlin 1998)

McGee and McLaughlin (1998) argue that Williamson's disquotational principles for truth presuppose the disquotational theory of truth. On the disquotational theory of truth, every instance of the Tarskian T-Schema is true. However, the correspondence theory of truth is compatible with the

rejection of universal validity of the T-Schema. Therefore, Williamson presupposes a particular theory of truth; his argument is not compelling if we opt for a different theory.

The problem with this objection is that Williamson's argument does not rely on instances of the T-Schema (Williamson 2004). It relies on principles that Williamson dubs 'disquotational'; however, this does not mean that the principles are only true on the disquotational theory of truth. Williamson's disquotational principles are substantively different from instances of the T-Schema. Instances of the T-Schema are contingently true; Williamson's disquotational principles are necessary. Semantic paradoxes like the Liar may require us to restrict the T-Schema to particular cases; for instance, we might want to restrict the T-Schema to apply only to sentences that manage to express a proposition (or to say something). Such a restriction does not apply to Williamson's disquotational principles: if an utterance fails to say anything, then the antecedent of the principle will be false, which makes the whole conditional true. Therefore, the objection seems to rely on a misunderstanding of Williamson's disquotational principles.

3.2.2 Williamson's disquotational principles are incorrect (Machina & Deutsch 2002)

Machina and Deutsch (2002) argue that the way that Williamson (1994b) presents his disquotational principles is not neutral. Williamson argues as if rejecting bivalence would lead to a contradiction: one sentence would have to be both true and false. However, the reason is that the disquotational principles are not stated neutrally; in neutral terms the disquotational principles should be stated as follows:

(Full Truth) If u says that p, then u is fully true if and only if p.
(Full Falsity) If u says that p, then u is fully false if and only if not p.

This is a neutral statement of the principles, because it leaves open the possibility of a sentence having some intermediate value between full truth and full falsity, which the original disquotational principles do not. Machina and Deutsch argue that if u is a borderline case, then u will be neither fully true not fully false. Therefore, the revised principles can be accepted but they by no means establish sharp boundaries for vague terms.

Williamson's (2002) reply to the argument that the epistemicist can interpret the talk of 'full truth and falsity' in epistemic terms (as definite truth and definite falsity understood epistemically). However, the revised principles are false on such a reading. If Michael is a borderline case of 'tallness' then if u says that Michael is tall, then it is not the case that u is definitely true if and only if Michael is tall. Furthermore, the epistemicist can accept that there are sentences that are neither fully true nor fully false, if being 'fully true' and 'fully false' are interpreted epistemically. Therefore, the revised principles are substantively different (and false) and not merely a neutral restatement of Williamson's original principles.

4 Objections to (2)

The second class of objections against epistemicist metasemantics is meant to show that even if the boundaries of vague expressions are sharp like the epistemicist claims, they are not shifty like the epistemicist claims they are. There are two lines of attack. The first is to argue directly against (2); the second is to undermine the epistemicist arguments for (2). I will consider these options in turn.

4.1 Direct objections against (2)

Anna Mahtani (2004) provides an interesting argument against (2). Williamson presents his argument for (2) as a dilemma. Given sharp boundaries for vague terms it is either the case that (i) the change in meaning occurs smoothly throughout the logical space or (ii) the change occurs in large leaps. Mahtani (2004) argues that Williamson's dilemma should be accepted. However, she believes that given Williamson's dilemma, (ii) should be accepted instead of (i). The reason is that the best explanation for why the boundaries of vague terms would be sharp in the first place is that each vague expression has a natural boundary. If vague expressions have no natural boundaries, it is difficult to see how they could be sharply bounded; since they are sharply bounded, it is most likely that they have natural boundaries.[40]

The first point to make here is that for certain vague expressions, it is completely implausible that some candidates for their reference would be more natural than others. Consider an expression like 'being pretty close': there are multiple distance properties that are candidates for its reference. It is widely accepted that all distance properties are equally natural. However, Williamson's arguments for sharp boundaries in case of such expressions works equally well.

Secondly, Mahtani's argument relies on the thesis that it is unlikely that usage, as a meaning determining factor, would fix a sharp boundary for vague

[40] Similar arguments lead Ross Cameron (2010) to propose a reference-magnetic version of epistemicism. If, as Mahtani (2004) argues, the most likely reason for why vague expressions have sharp boundaries is the fact that they have natural reference magnets, then in order to have the volatility in meaning required by epistemicism, we should claim that it is highly contingent which properties are more natural than others. Suppose that the actual cut-off point for baldness is 1000 hairs. Cameron's claim is that this boundary is actually more natural than other candidates; but it could easily be the case that some other candidate for the reference of 'bald' would be more natural.

terms. However, this is not the right way to look at the epistemicist arguments for (1) and (2). The epistemicist establishes sharp boundaries for vague expressions by appealing to global principles, like the disquotational principles for truth. Furthermore, the point of the Argument from Parity is that volatility in meaning is more plausible even if we don't make any assumptions about metasemantics. Even if one finds the epistemicist usage-based model implausible, there is an initial bias in favour of volatility of sharp boundaries.

4.2 Objections to the Argument from Use Metasemantics

Arguably, the largest amount of space in the literature on epistemicist metasemantics has been devoted to undermining the use-based metasemantic model presented by Williamson (1994b). I will discuss the objections posed in the literature below.

4.2.1 Incompleteness of epistemicist metasemantics (Horgan 1997)

The epistemicist presents a use-based metasemantic model. The epistemicist predicts that usage is sufficient to fix a sharp boundary for all vague expressions. However, the epistemicist does not explain the exact mechanism by which usage determines the sharp boundaries. The epistemicist does not present an account of metasemantic laws that would show how a sharp boundary is determined in a given case. This makes the epistemicist metasemantic model lacking.[41]

[41] A version of this challenge is presented by multiple authors. These include Burgess (2001), Horgan (1997), Horwich (1997), Keefe (2000, Ch. 3), Tye (1997), Weatherson (2003). The challenge is often presented in terms of partially defined expressions (expressions defined by incomplete stipulations); whereas I think that the case of incomplete stipulations is quite different form the case of vagueness (incomplete stipulations don't admit higher-order indeterminacy), some of the problems for incomplete stipulations have their analogues in the case of vagueness.

Williamson (1997b) argues that even if Horgan's argument were correct, all it would show is that the toy metasemantic model presented by the epistemicist is incomplete. However, this is not a big problem for epistemicism, because (1) and (2) are established independently of the metasemantic model. The metasemantic model provides a nice illustration of how (1) and (2) could be true, but epistemicism does not fall with the toy metasemantic model.

Secondly, note that epistemicism is not unique in relying on metasemantic generalisations (such as the thesis that usage facts determine a sharp boundary for vague expressions). Other theories like supervaluationism also rely on metasemantic generalisations, just different ones, e.g. supervaluationism requires that there are multiple sharp boundaries consistent with usage. Moreover, as argued above, such metasemantic generalisations are not problematic because their truth can be established by appealing to some global features (like the principles guiding truth) rather than local features of metasemantic laws.

In different work, Williamson (1996) provides a more direct arguments in favour of his metasemantic model. Suppose that Michael is a borderline case of tallness, but David is definitely tall. Presumably, the critics of epistemicism would concede that there is nothing problematic in classifying David as 'tall'. This is something that all the theories of vagueness agree on: somehow the usage facts (as well as potential other metasemantic factors) guarantee that David is in the extension of 'tall'. However, the epistemicist can reply that the same factors that guarantee David's inclusion in the extension of 'tall' can guarantee that Michael is in the extension of 'tall'. The difference between David and Michael is that we might be in a position to know that the metasemantic factors guarantee David's inclusion in the extension and not be in a position to know that they do so for Michael. But we shouldn't expect to know exactly where the metasemantic factors like usage draw the boundary,

since our knowledge of metasemantics is incomplete. The critics of epistemicism also need to draw a boundary between precisifications that are excluded by usage (and other factors) and the ones that are compatible with it. The epistemicist can make parasitic usage of this strategy: whatever fixes the boundary between definite cases and the rest also fixes the boundary between borderline truth and falsity.

This challenge to epistemicism is sometimes phrased in a slightly different way. Suppose that someone whose height is 180 cm counts as 'tall', but someone who is 179 cm does not. The question for the epistemicist is to explain why the first person and not the second one is in the extension of 'tall'. Williamson's standard reply is to say that our use is sufficiently liberal to admit those who are 180 cm in height into the extension of 'tall' but not sufficiently liberal to admit those who are 179 cm in height.

However, this reply seems unsatisfactory to some. The illustration of why this is the case will involve incomplete stipulations, but that is not essential to the problem. Suppose that the meaning of 'nice' is stipulated to meet the following criteria:

For all N>15, N is nice.

For all N<13, N is not nice.

Then there are three candidate cut-offs: 13, 14, and 15. It seems that they are on a par with regards how they are treated by the stipulation. (We can easily translate such a case to a case of a vague predicate: all we have to assume is that 13, 14, and 15 are such that no one ever used the predicate 'nice' to say anything about them, so there is no relevant use that could select one of them as one that fits best the pattern of use.) Therefore, it seems that they should all 'go together: either nice is true of all of them or false of them. Williamson (1994a; 1994b; 1997b; 1997c) argues that a case can be decided by *a default principle*, such as:

(Default Truth) For any atomic sentence S, if everything is equal between truth and falsity S is true.

(Default Falsity) For any atomic sentence S, if everything is equal between truth and falsity S is false.

Williamson's preference is for Default Falsity, which has some intuitive appeal; for instance, we naturally would expect that the above stipulation of the meaning of 'nice' does not support an interpretation on which Alpha Centauri counts as nice. So, it seems that we have a solution: default principles can decide the case for us.

One challenge to Williamson's use of default principles comes from Rosanna Keefe (2000, Chapter 3). Keefe claims that if Williamson has to rely on the default principles to solve such cases, then it means that his initial usage-based metasemantic model is insufficient. It is not enough to say that meaning supervenes on use to fix a sharp boundary. After all, there are many functions that would be compatible with meaning supervening on use: all that is needed for supervenience to hold is that a given pattern of usage is assigned the same interpretation at every world. If Williamson has to help himself to the default principles, it means that his usage-based model is insufficient.

The issue with this objection is that Williamson's default principles are just restatements of his original position. Recall Williamson-style explanation for why someone who is 180 cm in height counts as tall and someone who is 179 cm does not. The reason is that our usage is sufficiently liberal to admit the former person but not sufficiently liberal to admit the latter. But this is just equivalent to Default Falsity: unless usage is liberal enough to admit x into the extension of F, x is not in the extension of F. Therefore, the default principle is not an addition to the model, but merely a clearer restatement of the original position.

A more challenging objection to epistemicism comes from applying the default principles in more complex cases. Let's consider a case where instead of one predicate we have two: 'orange' and 'red'. Suppose that there is a standard series of patches from clearly orange on the left to clearly red on the right. Default Falsity would suggest that only those patches on the left that are sufficiently orange count as orange: the patches to the right are not orange. Similarly for 'red': the patches that are red are sufficiently red to count as red; the rest are not red. But it seems that there are patches that are neither sufficiently red to count as red nor sufficiently orange to count as orange. Nevertheless, we would like to say that all the patches are either red or orange. Default Falsity seems to be of no help in such cases.[42]

This objection relies on the assumption that every patch is either red or orange. However, it seems that the epistemicist has a plausible line of response once that assumption is also available to them. Suppose that it is a part of our use of language that in the above case every patch is either red or orange. Then the patches in the extension of 'red' will be ones in the anti-extension of 'orange' and vice versa: the patches in the extension of 'orange' are ones in the anti-extension of 'red'. The epistemicist would have a problem had the determination of extension and the determination of anti-extension been somehow different tasks. They are not. Setting the boundary for one also sets the boundary for the other. So it seems that the epistemicist has at least an initially plausible line of response to the problem.

[42] Brian Weatherson (2003) points out that the epistemicist can reply that for all that the case shows there is simply another colour between red and orange. However, this reply is unsatisfactory, because due to higher order vagueness the epistemicist would need to claim that there are indefinitely many colours between red and orange.

4.2.2 Truth*

An interesting challenge to the epistemicist metasemantics is presented by Keefe (2000, Chapter 3). Epistemicism requires that every vague predicate draws a sharp boundary between things in its extension and things in its anti-extension. For example, for some number N, N is the smallest number of dollars that one can have and be rich; one cent less, and one is not rich. Thus, the boundary between things that fall in the extension of 'rich' and those that fall in the anti-extension of 'rich' is drawn at N. Why should we take it that N is the cut-off point for 'rich'? The answer that this is because things that have N dollars or more are in the extension of 'rich', and those that do not, are not. But the objection goes, why should we care about the notion of *extension* rather than about a very similar notion of *extension**, which differs from *extension* only in that it is slightly more demanding: only things that have N+1 dollars are in the *extension** of 'rich'. Of course, saying that we should care about extension and not extension* is because all and only those things that 'rich' is true of are in the extension of 'rich', but are not in the extension* of it. The reply from the objector is to concede that it is not the case that all and only things that 'rich' is true of are in the extension* of 'rich', but there is a very similar notion to *truth*, namely *truth**, and all and only things that 'rich' is true* of fall in the extension* of 'rich'. Therefore, we have multiplied the notions, but have not answered the original question: why should we care about extension and truth rather than about extension* and truth*? This objection is discussed and solved in Chapter 6.

4.2.3 Radical externalism

The epistemicists need their metasemantic model to achieve two results. Firstly, the intensions of vague expressions have to be volatile across the space of nearby worlds. Secondly, in order for the epistemicist claim that we are not

in a position to know where the sharp boundaries are to hold any water, the metasemantic model cannot make it easy for us to figure out where these boundaries lie. The standard epistemicist explanation is to say that the sharp boundaries in our language are determined collectively by usage including the overall pattern of assent and dissent within the linguistic community.

However, some critics found it implausible that contents of our utterances and thoughts are determined by the use of expressions and concepts by other people.[43] The lesson from externalist arguments made by Putnam and Burge is that there are concepts and expressions such as 'water' or 'elm' the intensions of which are determined socially. In our use of such terms we defer to experts who are to tell us exactly what these expressions and concepts refer to. But there are no experts that would tell us where the boundary between the bald and not bald people lies. There are no 'baldness experts'. Therefore, there is no reason to suspect that we defer to other people in our use of concepts and expressions such as 'bald'.

It is often assumed that epistemicism is committed to radical social externalism, i.e. the view that all contents of our utterances and thoughts is determined by the use of concepts and expressions that compose our utterances and thoughts within our linguistic (or conceptual, if there is such a thing) community. The thought is that the epistemicist requires such a commitment for the epistemicist explanation of ignorance due to vagueness to go through. For example, it is possible for me to be extremely sensitive to the way that I use a certain vague word. If this is the case, then it may not be the case that I could easily use the word even slightly differently. So if the meaning of the word in my mouth would have been determined solely by how I use the word, then there would be no close possible world where the meaning of the word

[43] See Keefe (2000, Chapter 3) and Schiffer (1999).

shifts, so the word would not be semantically plastic; however, it would still be vague, so semantic plasticity would be independent phenomenon from vagueness and vagueness would remain unexplained.

It is in fact a misreading of the theory to claim that epistemicism is committed to radical social externalism (Williamson 2002). Epistemicism is not committed to the impossibility of a private language. For example, it is compatible with someone introducing a vague word intro their vocabulary and using it privately. However, the epistemicist insists that if the word is vague, then one does not know where the sharp boundaries of the word are. But insisting that our knowledge of the meanings of words (even if this is restricted only to private words) entails that we are in a position to decide every case can only be justified on very implausible verificationist assumptions, e.g. that to know what the word means is to be in a position to know whether the word applies in every case. Such assumptions are implausible, because we are not in a position to know exactly what our dispositions to use expressions and concepts are. For example, a change in my disposition to use a word in one context does not entail that the disposition changes in another context; the effects of changes in my dispositions to use a particular word may not be easily traceable.

Furthermore, it is entirely possible that someone decides to be very careful about the use of their private word. For instance, someone may choose to apply the word 'ptall' similarly to how we use the word 'tall', but make sure to apply it to men who are over 180 cm tall. It is very plausible that the word 'ptall' in their mouth picks out the property of having over 180 cm and this is known to them. However, this situation is akin to a case where one simply stipulates that they will use the word 'tall' to refer to people that are over 180 cm in height. Even if the stipulation is successful, this does not mean that the question about our old and standard way of using the word 'tall' has been

answered; the stipulation only makes the word 'tall' have a different meaning and says nothing about the old meaning of the words. On the new stipulated meaning, the word 'tall' is not vague in any relevant sense. In such a situation there are no relevant borderline cases for the epistemic view to explain, so such cases pose no threat to epistemicism.

Williamson often invokes models of communication between speakers which suggest that he endorses radical semantic externalism. The reason for this is quite simple. If it were the case that the speaker's use of a vague word were the sole determinant of the meaning of the word in their mouth, then communication failures would be extremely common.[44] It is practically guaranteed that in any given communication scenario, the speaker and the hearer would not have identical dispositions in using words that they employ in the conversation. Suppose that A and B differ in their dispositions to use the term 'bald' so that in A's mouth 'bald' denotes baldness_1 and in B's mouth it denotes baldness_2. Then, if the speaker A says 'X is bald' and the hearer Y responds 'X is not bald', it may very well be the case that they are speaking past each other: A expresses some proposition p (that X is bald_1) and B expresses a proposition that is not a negation of p, but of some other proposition (that X is bald_2).

However, it is clear that we don't experience communication failures that frequently between members of the same linguistic community and that there are many instances of successful communication. In cases of successful communication, it is reasonable to assume that the meaning of words that are employed by the speaker and the hearer is parasitic on both of their uses. If

[44] Schiffer (1999) raises this issue as a problem for epistemicism. However, the fact that speakers within a linguistic community differ with regards to their semantic intensions and disposition is something that everyone has to contend with. It is not a special problem for epistemicism.

this is the case, then the meaning of a word in one speaker's mouth does not depend solely on her dispositions to use the word but depends also on the dispositions of other speakers. According to the epistemicist such models of communication make the epistemic view of vagueness more plausible. Coming back to the case from the beginning of this section, one may be very careful in how one uses a particular word, but the meaning of the word in one's mouth may still shift because of some shift in the disposition to use the word by a different speaker. So it seems that in order to explain successful communication, we have to externalise the meaning in speakers mouths to a certain extent; this makes it more plausible that we are not perfectly attuned to small shifts in the meanings of the words we use.

The epistemicist point is also that this is supposed to be only a toy model for how communication works. It seems that no one has a fully worked-out theory of communication that would reconcile differences in dispositions to apply the words by different speakers with successful communication. However, if such a theory arises, it is likely to go along the lines outlined by Williamson. Most importantly, there is no evidence that alternatives to epistemicism have a simple solution to these issues. Everyone needs a good theory of linguistic communication; other theories of vagueness do not give us an easy route to providing such a theory. There does not seem to be a well worked-out theory of communication that would not be consistent with vagueness. Requiring the epistemicist to provide a working theory of communication would be to put too much burden on epistemicism in contrast with other theories of vagueness.

4.2.4 Heavyweight vocabulary and moral vagueness

The most challenging problem for epistemicist metasemantics, in my opinion, has been raised by Miriam Schoenfield (2016). Schoenfield argues that epistemicist metasemantics may be suitable for lightweight vocabulary like

'tall' or 'bald'. It makes sense that usage may determine which candidate for the reference of these terms, since it is not really important which particular candidate is selected (nothing much hangs on the facts whether the cut-off point for baldness is 1000 as opposed to 1001 hairs). However, that picture is not plausible in case of more heavyweight vocabulary. Schoenfield argues that the epistemicist metasemantic model is unsuitable for dealing with vagueness of moral vocabulary such as 'permissible'.

However, the challenge points to a more general weakness with the epistemicist model that extends to non-moral pieces of vocabulary as well.[45] For instance, presumably the term 'water' is vague. 'Water' does not always refer to pure H_2O, but to various mixtures of H_2O and other substances; it is vague how much other substances can be added to H_2O to still count as 'water'. Nevertheless, the Putnamian (1975) orthodoxy has it that the properties that causally regulate the way we use terms like 'water' also play an important role in determining its reference.

The initial response from the epistemicist is that the use-based metasemantic toy model is not very refined. At the outset of Chapter 1, I outlined two ways of understanding usage. Examples of moral vocabulary and natural kind terms, indicate that for epistemicist model to be plausible, we have to understand usage broadly. However, the epistemicist metasemantic model is not very refined. 'Usage' is treated as a catch-all term: to say that meaning supervenes on usage is not to say much more than to say that meaning supervenes on whatever it supervenes on. I think the reason why the details of the use-based model where not developed further, apart from the fact that we

[45] I do not claim that the cases of moral vocabulary and natural kind terms are identical. As demonstrated by the Moral Twin Earth cases, they are not. Nevertheless, both cases demonstrate that there is a weakness with the metasemantic toy model employed by the epistemicist, by showing that usage is not everything.

don't know much about metasemantics in general, is very prosaic. The majority of the discussion on vagueness concerns lightweight expressions like 'bald' and 'tall'. Non-epistemicist philosophers had trouble imagining how it would be possible for lightweight vague expressions like 'tall' or 'bald' to have sharp and shifty boundaries. The use-based model was meant to show how that would be possible: small shifts in usage facts produce small shifts in reference. The details of what constituted usage were not developed, since in cases of such lightweight vocabulary it can be assumed that usage reduces to narrowly construed usage (patterns of assent and dissent). However, once the debate progresses, epistemicists have to present a more detailed metasemantic account that would allow epistemicism to handle vagueness of more heavyweight vocabulary. My aim in Chapter 3 is to achieve part of this task: I will present a version of epistemicist metasemantics that can handle the vagueness of moral terms.

5 Conclusion

In this chapter, I outlined the most common challenges to epistemicist metasemantics. I also presented what, I think, are the best lines of response for the epistemicist to address these challenges. The discussion is by no means conclusive: the best I could hope to achieve in this chapter is to show that there is no immediate knockdown punch against the epistemicist metasemantics. The epistemicist does have some defensive resources. However, there is an issue, I think, that should be addressed in greater detail: the argument against epistemicist metasemantics from moral vagueness (Schoenfield 2016). Answering this challenge is the goal of the next chapter.

Chapter 3

Shifty Morals

ABSTRACT

Epistemicism explains ignorance due to vagueness through semantic plasticity: the propensity of intensions of vague terms to shift across close linguistic communities. In the case of moral vagueness, e.g. when it's vague whether it's permissible to terminate a pregnancy after a certain number of weeks, epistemicism predicts that 'permissible' denotes distinct properties in different close linguistic communities. This epistemicist prediction has been recently pressured by Miriam Schoenfield (2016) and Robbie Williams (2018). Firstly, Schoenfield (2016) argues that if there is moral vagueness, it is ontic. Secondly, in response to the problems posed by Moral Twin Earth cases, Williams (2018) presents a metasemantic theory which predicts that moral terms are not semantically plastic. In this chapter, I address these challenges by stating an epistemicist-friendly metasemantic theory of moral terms that is compatible with the Moral Twin Earth intuitions and avoids the challenges posed by Schoenfield (2016). I make use of the metasemantic framework for moral terms given by Robbie Williams (2018), and I refine it to give an epistemicist account of moral vagueness.

WORD COUNT: 8,756

1 Introduction

Suppose that Darryl is watching his two-year-old daughter playing in the park. Presumably it is permissible for him to look away for 5 seconds, but impermissible to look away for 5 minutes. However, we cannot point to a precise boundary between permissible and impermissible actions in this situation: we don't know the longest period for which it is permissible to look away. Thus, 'permissible' seems to be a vague term.

According to the epistemicist theory of vagueness, vague terms are semantically plastic, i.e. they are characterised by easy shifts in reference across close possible worlds. In Darryl's case, the epistemicist theory predicts that we are not in a position to know the longest period that it's permissible to look away for because 'permissible' could easily denote another property with a slightly different cut-off. Thus, any belief regarding the location of the cut-off is unsafe: we could easily make an error due to the shiftiness in reference of the vague term in question.

The epistemicist treatment of moral vagueness, in particular the thesis that moral terms are 'shifty' as required by the epistemicist theory, has recently come under pressure. Miriam Schoenfield (2016) argues that moral vagueness is incompatible with epistemicism. The basic idea behind Schoenfield's objections is that shiftiness in reference may be plausible for terms such as 'tall', but not for normatively significant words such as 'permissible'.

A related worry comes from the Moral Twin Earth cases, which produce the intuition that the reference of moral terms such as 'permissible' remains stable across different linguistic communities. In a Moral Twin Earth case, two societies: Kantopia and Utilitas, use their moral terminology differently; the citizens of Kantopia use their moral terminology in accordance with the prescriptions of Kantian ethics while inhabitants of Utilitas use it to accord

with Utilitarian prescriptions. On a metasemantic model on which the reference of a term is determined by its usage in a linguistic community, the reference of moral terminology is distinct in both communities: in Kantopia it refers to Kantian properties and in Utilitas to utilitarian ones. However, our intuitions in Moral Twin Earth cases tell us that despite differences in opinions, the reference of moral terminology should be the same for both communities: it's not enough to embrace a particular moral theory to change the meaning of moral terminology. The problem for epistemicism is that metasemantic models that are meant to account for the Moral Twin Earth intuitions, such as the models proposed by Robbie Williams (2018) or Billy Dunaway & Tristram McPherson (2016), predict that moral vocabulary is stable, which makes them incompatible with epistemicism.

My main aim in this chapter is to present a version of the epistemicist theory that is compatible with moral vagueness.[46] After some clarifying preliminary comments, I outline the epistemicist theory and the problem posed by moral vagueness and the Moral Twin Earth intuitions. Then, I sketch the view provided by Williams (2018; 2020) and I show that it needs refinement to handle moral vagueness. I present two versions of the refined account and I show that one is better at handling higher-order vagueness. Lastly, I solve remaining problems for the epistemicist treatment of moral vagueness, mostly due to Schoenfield (2016), and I show how they can be addressed by the epistemicist.

[46] See Manley (MS) for an epistemicist account of moral vagueness that rejects the Moral Twin Earth intuitions.

2 Preliminaries

2.1 Epistemicism

Epistemicism attempts to explain ignorance due to vagueness by semantic plasticity:

> (Semantic Plasticity) An expression E is semantically plastic if and only if the shifts in the usage of language across close possible worlds imperceptibly shift the intension of that expression.

The epistemicist wants to ultimately show that semantic plasticity leads to ignorance. Use of language shifts easily across possible worlds. Vague words are semantically plastic, so the intensions of these words shift across close possible worlds. Thus, in borderline cases a sentence such as 'John is bald' may be true in the actual world and false in some close possible world, due to the shift in the intension of 'bald'. Furthermore, the epistemicist claims that this phenomenon accounts for ignorance due to vagueness, because beliefs expressed by borderline sentences can easily be false (and knowledge requires safe beliefs).

2.2 Moral realism

Miriam Schoenfield (2016) argues that for a moral realist, epistemicism is incompatible with moral vagueness. Schoenfield (2016) understands moral realism as the position that moral properties such as *permissibility* are fundamental: if there was a fundamental language (such as the language proposed by Sider (2011)), then moral vocabulary would be a part of that language. I will follow that characterization of moral realism. For convenience, I will use the term 'fundamental' to talk about both properties (*permissibility)* or expressions ('permissible'): an expression is fundamental if it's part of the fundamental language.

2.3 Moral vagueness

I am going to focus on moral vagueness in this chapter, so it's worth to briefly clarify what that is.

Firstly, I assume that there is moral vagueness in this chapter. For instance, I assume that part of the reason why we don't know whether for some number N it is permissible for Darryl to look away for N seconds from watching his daughter play in the park is that 'permissible' is a vague term. That assumption may be incorrect. Perhaps the reason for our ignorance is that our moral theories are still underdeveloped. Furthermore, the exact cut-off point for permissibility may be context-sensitive and determined by multiple factors: in Darryl's case this could be things like the number of people in the park, the time of day, the crime rate in the area etc. Doing such multifactor analyses is difficult, so it would be no wonder that we don't know where the cut-off point for permissibility is. For the purposes of this chapter, I will assume that there is moral vagueness and Darryl's case is a paradigm example of it.

Secondly, I assume that if a term is vague, then there are multiple candidates (or precisifications) for its reference. For instance, if 'permissible' is vague, then there are multiple candidates for the reference of 'permissible'. What it means to be a 'candidate' depends on the theory of vagueness one embraces. For instance, on semantic theories of vagueness, a candidate for reference of 'permissible' would be any property that is compatible with how 'permissible' is used in the language. On the other hand, on ontic theories of vagueness, there are multiple candidate properties such that the world is unsettled as to which one is *permissibility*.

Thirdly, I will focus only on borderline cases that arise as a result of vagueness of moral vocabulary such as 'good' or 'permissible' such as in Darryl's case. In Robbie Williams's (2016) terms, I will focus on normative

vagueness, not factual vagueness.[47] To illustrate the distinction, suppose that I promise you a blue rock and I give you a rock that is a borderline case of blue. Thus, it is borderline whether my giving you the borderline blue rock is permissible (to satisfy the promise). But the vagueness is inherited from the vagueness of 'blue' in this case.[48]

Fourthly, I will focus on borderline cases of moral vocabulary such as 'permissible' or 'wrong' assuming that they are fundamental moral terms by the lights of our first-order moral theory. Not every first order moral theory takes the same moral terms to be basic or fundamental. This matters because focusing on non-fundamental moral terms may be less tricky as opposed to the case of fundamental moral vocabulary. Here is an example. Scalar Utilitarianism (Norcross 2006) takes 'better than' and not 'permissible' to be fundamental. There is an easy way of combining an epistemicist account of vagueness for terms like 'permissible' with Scalar Utilitarianism. When we established by moral deliberation which ordering is correct, we have settled all *fundamental* moral facts that there are to settle. Therefore, there will be no problem in accounting for the vagueness of 'permissible' when 'permissible' is not a fundamental term by the lights of the first-order moral theory. That is not to say that issues related to moral vagueness will not arise on Scalar Utilitarianism (though this is what Schoenfield 2016 seems to assume). It is possible that issues related to moral vagueness will arise on Scalar Utilitarianism, because it will be vague which ordering on actions is the correct one (e.g. as a result of incommensurability cases[49]). My point is merely that it

[47] This is not to suggest that there are no normative facts.

[48] Similarly, a sentence like 'John is over 176 cm' may be vague, because it may be vague what are John's spatial boundaries, does his hair count towards his height etc. This does not mean that 'is over 176 cm' is a vague term – the vagueness of the sentence is inherited form the vagueness of 'John'.

[49] See Chang (1998; 2002).

is easier to account for vagueness of a moral term, if we adopt a first order moral theory that does not take that particular term to be fundamental. To avoid such cheap tricks, when I focus on a vagueness of a moral term my intention is to treat this term as fundamental by the lights of our first-order moral theory.

In the next section, I will outline the problems for epistemicist account of moral vagueness.

3 Problems for epistemicism

There are two lines of argumentation that pose a problem for the epistemicist treatment of moral vagueness. First, the Moral Twin Earth cases seem to predict that moral terms are stable in a way that is incompatible with the shiftiness of vague terms required by epistemicism. Second, Miriam Schoenfield (2016) explicitly argues that moral vagueness is incompatible with epistemicism. I will outline these problems and then show that there is a common thread between them.

3.1 Moral Twin Earth

Suppose that two communities: Kantopia and Utilitas use the term 'permissible' in different ways. The community on Kantopia embraces Kantian ethics and applies the term 'permissible' accordingly. On the other hand, the community on Utilitas is utilitarian and use 'permissible' in ways prescribed by utilitarianism. Suppose that we have a simple metasemantic model on which the only factor determining the reference of 'permissible' is the usage of the term within a linguistic community. The reference of 'permissible' would be whatever property best fits with usage and maximises the number of true utterances within the community. On such a model, 'permissible' would have

different reference on Kantopia and Utilitas since the usage is different in both communities: on Kantopia it favours the Kantian notion of permissibility and on Utilitas the utilitarian one.[50] This is a problem for a few reasons. Firstly, it is not clear how the two communities could communicate with one another. When a citizen of Kantopia utters 'A is permissible' and the inhabitant of Utilitas utters 'A is impermissible' they are not really in disagreement – they are merely talking past one another. Secondly, as Horgan & Timmons (1991) claim in presenting the Moral Twin Earth cases, there is a powerful intuition that despite the philosophical disagreement between the communities over which moral theory (Kantianism vs. Utilitarianism) is correct, the reference of 'permissible' would be the same for both communities.

Moral Twin Earth poses a problem for the epistemicist theory of vagueness, because it presents a powerful case that reference of expressions like 'permissible' should remain stable throughout significant changes in usage.[51] In contrast, epistemicism requires the reference of vague expressions to be shifty.

---

[50] The point of MTE cases was to reject even more sophisticated models for the metasemantics of moral terms. Suppose that we allow for the external worldly facts to play a part in determination of the reference of moral terms, e.g. if we allow that the property that causally regulates the use of the term 'permissible' is its referent. MTE cases show that this will not do, since it still seems that the properties that causally regulate the use of 'permissible' (and thus should be the referents of 'permissible' according to Putnam's externalism) in Utilitas and Kantopia are distinct.

[51] It's worth to point out the contrast between this point and Williamson's argument from Parity explored in Chapter 2. Williamson's argument relies on the claim that it is implausible to think that meaning of vague terms like 'bald' shifts only as a result of large shifts in usage. It is much more likely that small changes in usage produce small changes in meaning. The point of the Moral Twin Earth cases is that not even very large shifts in usage may be able to shift the meaning of moral terms. Communities that have a very different usage of moral terms (e.g. as a result of endorsing a very different moral theory) still refer to the same moral properties that we do, as long as the moral terms they use play a similar role of regulating behaviour as they do for us.

Thus, there is a tension between Moral Twin Earth intuitions and semantic plasticity that the epistemicist should resolve.

3.2 Moral vagueness as a problem for epistemicism

The second line of argumentation against epistemicism is due to Miriam Schoenfield (2016). Suppose that it is borderline whether it is permissible to terminate a pregnancy after 150 days. Cheryl is 150 days pregnant and is unsure whether it is permissible for her to terminate the pregnancy. In order to find out whether it is permissible, she consults a linguist who tells her how the word 'permissible' is used within her linguistic community (alternatively: she consults an anthropologist to find out what conceptual role 'permissible' plays within the community). On the epistemicist picture, this would be a good way for Cheryl to proceed as what counts as 'permissible' depends on how the word is used within the linguistic community. However, this is not a reasonable view of moral deliberation: one does not learn moral truths by learning anthropological truths about a given community. To answer the question, it is not enough for Cheryl to learn facts about linguistic usage. So epistemicism has implausible consequences for moral deliberation.

In the next section I will show that both these lines of argumentation contain a common thread.

4 Towards an epistemicist solution

Let's first consider Schoenfield's challenge to epistemicism. Why is finding out what's permissible by learning facts about linguistic usage problematic? The first thing to clarify is that Schoenfield does not claim that epistemicism is committed to the view that facts about what's permissible are dependent on facts about linguistic usage. It is possible within the epistemicist framework

that w^*: this does not mean that ϕ is permissible in w^*, but rather that 'ϕ is permissible' expresses a different proposition in w^* that is true in w^*. So, the challenge is not that epistemicism is committed to moral facts being somehow dependent on facts about language or conceptual roles.

To appreciate the challenge, let's consider a case of non-moral vocabulary. Suppose that we would like to find out what property the word 'water' refers to. We would go to a linguist or an anthropologist who would tell us the story about what 'water' refers to. Assuming a Putnamian (1975) picture of how the reference of natural kind terms is determined, the anthropologist could tell us that the word 'water' is used within our linguistic community with the intention that it refers to a substance with a unified underlying chemical structure. He could also tell us that chemists have found out that the underlying chemical structure of water is H_2O. Furthermore, he could tell us that in everyday life 'water' is used as a vague term: we don't use 'water' to denote pure H_2O. Rather we use 'water' to denote mixtures of H_2O with minerals and other elements. Assuming that he is an idealised all-knowing anthropologist, he could tell us where exactly the boundary for 'water' is, e.g. how many drops of milk can we put in a glass of water before it ceases to be water.

Clearly, not all facts regarding the reference of 'water' that we would learn from the anthropologist would be facts about linguistic usage. Most notably, in order to find out what the reference of 'water' is, we must learn some substantive facts about the chemical composition of water. Therefore, we cannot straightforwardly learn what the reference of 'water' is from an anthropologist, his anthropological knowledge must be supplemented by the knowledge of the (substantive) findings of chemistry. Nevertheless, some of the anthropological or linguistic knowledge would also be relevant in determining

the reference of 'water' (how much non-watery stuff can we add to water and still have water in the glass).

Similarly in the moral case, the anthropologist can tell us some things about the usage of 'permissible' which will be useful in finding out the location of the boundary in Cheryl's case. However, this does not mean that there are no substantive moral facts that play a large role in determining the reference of 'permissible' just like substantive chemical facts played a large role in determining the reference of 'water'.

Schoenfield's challenge to epistemicism relies on an erroneous assumption about epistemicist metasemantics of moral terms.[52] The epistemicist is not committed to the claim that substantive moral facts do not determine the reference of moral terms. However, there remains a challenge for the epistemicist to provide an account of the metasemantics of moral terms that would allow for the reference of these terms to be determined by substantive moral facts. This would allow the epistemicist to make sense of moral deliberation: Cheryl could not simply rely on the anthropologist's knowledge in answering the question whether abortion in her case is permissible, she would also need to gain knowledge of substantive moral facts.

What about the challenge to epistemicism from the Moral Twin Earth cases? The Moral Twin Earth cases boost the intuitions that large differences in usage of moral terms (usage favouring Kantianism in Kantopia and

[52] Epistemicists are often pressured to explain the semantic plasticity of vague terms: how is it possible that their reference is so prone to shift (Mahtani 2004, Cameron 2010)? The standard epistemicist reply is to explain the shiftiness in reference by shiftiness of usage (Williamson 1994b). Linguistic communities across close possible worlds use language slightly differently, e.g. some communities use a term like 'bald' more liberally, so that 'bald' in their mouths applies more widely, which explains why 'bald' as they use it has a slightly different reference. The fact that epistemicists use this reply leads some to conclude that the epistemicist claim that (at least in the case of vague terms) narrowly construed usage is the only meaning determining factor. This is not the case: usage should be construed broadly.

utilitarianism in Utilitas) are insufficient to shift the reference of moral terms. If large shifts in narrowly construed usage between communities are insufficient for large shifts in reference of moral terms, how is it possible for small shifts in usage to produce small shifts in reference (as required by epistemicism)?

I think there are two issues to consider here. The first issue is related to how the reference of moral terms is determined. MTE cases show that the model, on which the reference of moral terms is determined by which candidate best fits narrowly construed usage, is wrong. On such a model it would be enough to change the use of an expression, like 'permissible', in order to change its reference. However, our moral intuitions (insofar as we are moral realists) say that this is not the case: it's insufficient to have a wrong moral theory (and thereby corresponding patterns of assent and dissent) to change the reference of 'permissible'. For instance, whole populations can be mistaken about some moral truths such as the moral status of slavery. [53] Even if most people within a population believe that slavery is permissible (and thus the sentence 'slavery is permissible' adheres to the narrowly construed usage within the community), it does not mean that 'permissible' as they use it refers to a property that allows for slavery. Fortunately for the epistemicist, as discussed in Chapters 1 and 2, this is not the epistemicist model. On the epistemicist model, usage is

[53] This requirement is not special to the case of moral vocabulary (though the metasemantic details may be slightly different). We also want to allow for a possibility of such systematic error, e.g. in case of natural kind terms like 'fish'. We once thought that whales are fish, not mammals. Since everyone thought that whales are fish, whales would fall in the extension of 'fish' if fitting most neatly with usage is the only factor that is taken into account in our metasemantic theory of 'fish'. However, what we want is a metasemantic theory that allows for systematic error: when we use the term 'fish' we intend to refer to a natural kind, in which case part of what determines the reference of 'fish' is e.g. the genetic composition of objects in its extension. Therefore, we want a metasemantic theory that allows for more reference-determining factors than narrowly construed usage for more general reasons than the specific moral case.

construed broadly and used as a kind of catch-all term. So the conclusion that epistemicist metasemantics is incompatible with Moral Twin Earth intuitions would be too quick. Nevertheless, epistemicism has to present a model that would allow for combining semantic plasticity with Moral Twin Earth intuitions that the reference of 'permissible' can withstand large variations in narrowly construed usage. Thus, there is a common challenge to epistemicism coming from the Moral Twin Earth cases and Schoenfield's objections.

The second issue is the tension between the epistemicist claim that vague terms (including vague moral terms) are semantically plastic and the Moral Twin Earth cases which boost the intuition that the reference of moral terms is stable across large shifts in narrowly construed usage. Fortunately, there is room for compromise. As I see the MTE cases, they pump the intuition that the disagreement between the two communities should not always be purely verbal, i.e. if a conceptual role of a term is similar (or the same) in two communities the reference of the term is similar as well. This relates to the first issue: in case of moral terms, we don't think that having a wrong moral theory (or using the moral terms in a non-standard or morally incorrect way) is sufficient to make moral terms refer to properties that would validate such use. But we should not give in to the intuition completely and claim that the reference of moral terms for the two MTE communities is identical. If we extrapolate to cases of other predicates, the MTE intuitions would yield the wrong results. For instance, 'tall' plays a very similar conceptual role in our community as in any close linguistic community; but we shouldn't automatically conclude that the reference of 'tall' is identical in all nearby linguistic communities. The reference of 'tall' can shift across close linguistic communities. So, if we took the MTE intuitions to require automatically that the reference is identical in the two communities because the conceptual role is, we would overgeneralize. What the MTE cases show is that given the

similarity of conceptual roles that moral terms play on Kantopia and Utilitas, moral terms like 'permissible' should refer to fundamental moral properties in both communities.

Therefore, the task for the epistemicist is to present a version of a semi-stable metasemantic theory for moral terms. The theory should not predict that the reference of moral terms is stable across shifts in usage (or conceptual roles): we need some shiftiness for the epistemicist account of ignorance due to vagueness to work. Furthermore, the theory should not predict that the reference of moral terms can freely shift with usage. In case of Kantopia and Utilitas, the theory should predict that at least one of these communities has a wrong moral theory. In the following sections, I will present a version of semi-rigid epistemicist theory that satisfies the Moral Twin Earth intuitions.

In the next section, I will outline Robbie Williams's (2018; 2020) metasemantic framework. The framework, as presented, predicts semantic stability which is a problem for the epistemicist. However, we will be able to adjust Williams's model and turn it into a semi-stable metasemantics for moral terms that's required by epistemicism.

5 The reason-responsiveness metasemantics for moral terms

Robbie Williams (2018; 2020) gives an account that allows for stabilising the reference of moral terms in Moral Twin Earth cases.[54] The account makes use of three theoretical tools: conceptual role determinism, substantive radical

[54] Robbie Williams (2018) presents his account as a theory of mental representation; I will focus on linguistic representation, but I am not distinguishing the two for the purposes of this chapter (if a linguistically competent agent uses ϕ to express his belief then the content of ϕ is identical to the content of his belief).

interpretation, and normative principles linking blame and reason-responsiveness. The three elements can be briefly outlined as follows:

> *Conceptual Role Determinism.* The basic idea behind conceptual role determinism is that the reference of a term supervenes on its conceptual role within a linguistic community. Suppose that a term W plays the conceptual role of assigning blame: (a) when the agent judges x's A-ing to be W, this makes them blame x for A-ing, (b) when the agent judges x's A-ing not to be W, this prevents them from blaming x for A-ing (Williams 2018). Necessarily, if W plays this role, then W picks out the property P. Let's assume with Williams (2018; 2020) that P is the property of violating the categorical imperative.
>
> *Substantive Radical Interpretation.* The radical interpretation approach to metasemantics is one that says that the correct interpretation of an agent's utterances is the one that best explains her dispositions. On Williams's version of radical interpretation, 'best explanation' is understood in terms of rationality: the correct interpretation of agent's utterances is the one that makes them most rational. One way of cashing out this notion is via structural rationality: agent's beliefs and desires should meet some basic structural constraints; for instance, a correct interpretation of an agent should not assign to them both the belief that *p* and the belief that *not p*. However, such structural constraints seem to be insufficient for narrowing down the correct interpretation of an agent (Williams 2020, p. 17-19). Therefore, Williams opts for a version of radical interpretation that supplements the structural rationality constraints with some *substantive* constraints: a correct interpretation is one that also does not attribute to the agent some weird beliefs and desires (e.g. such as the desire for a saucer of mud[55]). Williams argues that the best interpretation of the agent's utterances is one that would make her the most reason-responsive.
>
> *Normative principles.* The normative principles connect reason-responsiveness and wrongness to the effect that to be reason-responsive

[55] See Lewis (1983, p. 375; 1986, p. 38-39). Lewis takes the 'desire for a saucer of mud' example from Anscombe (1957, §37).

is to assign blame to all and only those actions that are wrong. They go as follows (Williams 2018):

(1) A reason-responsive agent would be such that: for any agent x and act A, the judgement that x's A-ing was wrong (and unexcused), makes them blame x for A-ing.

(2) A reason-responsive agent would be such that: for any x and A and for any F other than features that entail wrongness, the judgement that x's A-ing is F would not make them blame x for A-ing.

(3) A reason-responsive agent would be such that: for any x and A, the judgement that x's A-ing is not wrong, would make them refrain from blaming x for A-ing.

(4) A reason-responsive agent would be such that: for any x and A and for any G other than wrongness entails, a judgement that x's A-ing is not G would not make them refrain from blaming x for A-ing.

These three elements combined allow us to have semantic stability across different linguistic communities. In the Moral Twin Earth case, if wrongness is the property of violating the categorical imperative, then even for the consequentialist community their term 'wrong' refers to wrongness, because 'wrongness' plays the same conceptual role in the two communities and only under this interpretation we maximise their reason-responsiveness: only the fact that some action is wrong gives us a reason to blame someone for performing it.

Therefore, Williams's account gives us stability of moral terms and allows for the possibility of systematic error. However, as it stands the account does not seem to allow for semantic plasticity: it forces absolute stability of reference without sensitivity as to whether the terms in question are stable or plastic. The task in the next section will be to refine the framework to accommodate semantically plastic moral terms. This will allow us to give an epistemicist treatment of moral vagueness in later sections.

6 The epistemicist solution to moral vagueness

I can see two ways to adjust the reason-responsiveness approach to make it more flexible and allow it to deal with vagueness. I will call them Solution A and Solution B. I think both approaches are initially promising. However, I think Solution A faces a serious problem when it comes to dealing with higher-order vagueness. Solution B is much better equipped for this task. Therefore, I will opt for Solution B in the end.

6.1 Solution A

Solution A[56] relies on relaxing the Conceptual Role Determinism thesis and the four normative principles so that multiple candidates are allowed as referents of 'wrong'. The philosophical justification for this move is that in borderline cases it does not seem like assigning one as opposed to another candidate as the reference of one's utterances makes one more reason-responsive. In order to do this, we have to firstly relax Conceptual Role Determinism. Instead of allowing for there being only one candidate for the reference of W for any given conceptual role R, we allow that there are multiple candidates (call them wrongness-candidates) for the reference of W. The idea is that an interpretation of W as referring to one of the wrongness-candidates does not make an agent any more reason responsive than an interpretation of W as referring to another wrongness-candidate: multiple interpretations make the agent equally reason responsive.

Secondly, we have to adjust the normative principles, so that they allow for multiple wrongness-candidates. Let's say that an act is determinately wrong

[56] The position defended in Sud (2019) can be interpreted in this way. Sud (2019) argues for a semantic account of vagueness, but his framework can be easily adopted to epistemicism.

(not wrong) if it is wrong (not wrong) according to every wrongness-candidate. Our relaxed normative principles are given by:

(1*) A reason-responsive agent would be such that: for any agent x and act A, the judgement that x's A-ing was determinately wrong (and unexcused), makes them blame x for A-ing.

(2*) A reason-responsive agent would be such that: for any x and A and for any F (such that it is not the case that F entails any of the wrongness-candidates), the judgement that x's A-ing is F would not make them blame x for A-ing.

(3*) A reason-responsive agent would be such that: for any x and A, the judgement that x's A-ing is determinately not wrong, would make them refrain from blaming x for A-ing.

(4*) A reason-responsive agent would be such that: for any x and A and for any G (such that G is not entailed by every wrongness-candidate), a judgement that x's A-ing is not G would not make them refrain from blaming x for A-ing.

Importantly, what the revised principles tell us is that if an act is definitely wrong, then a rational agent would blame one for committing it and if an act is definitely not wrong then a rational agent would not blame one for performing it. But the principles give us leeway in borderline cases: we are neither required nor forbidden by the normative principles to blame someone for performing a borderline wrong act. On such an approach the rationality considerations are constraints on reference, but don't uniquely secure reference if there are multiple candidates for the reference of 'wrong' (i.e. if there is more than one property among the Ps).

The refined principles have an advantage over the initial version, because they don't force precision where it is not required. One problem that arises though has to do with the opposite worry that the principles may produce vagueness where we require precision. For one thing, it would be desirable to have metasemantics that would be capable of handling both vague and precise vocabulary. Fortunately, the refined principles don't require vagueness nor precision: the principles work equally well when there is one or multiple candidates for the reference of a term (in our case 'wrongness'); in case where there is only one admissible candidate, the refined principles are equivalent to the unrefined ones.

Here is the picture so far. For the two Twin Earth communities the use of the term 'wrong' plays the same coarse-grained conceptual role. This means that despite different beliefs about the subject matter (what is wrong, what is permissible) the reference of the crucial terms like 'permissible' or 'wrong' is roughly the same for both communities. This means, for instance, that if Kantianism is true, both communities are interpreted as referring to a Kantian property. This much we get from the refined Williamsian metasemantics. However, the relaxed principles allow for there to be multiple candidates for the reference of moral terms like 'wrong': they impose constraints on reference, but don't force stability of reference. Williams's model allows us to get the desired semi-stability. However, the relaxed principles admit multiple candidates for the reference of moral terms (multiple interpretations of agents make them equally rational). There remains a matter of accounting for the semantic plasticity of moral terms. Epistemicism requires that different close linguistic communities refer to slightly different properties in case of vague terms. Therefore, we can let narrowly construed usage decide which particular candidate (among the ones that satisfy the refined normative principles) is assigned as the reference of moral terms for a given linguistic community.

Unfortunately, one important problem arises for the account: higher-order vagueness. With Solution A we managed to explain first-order moral vagueness (i.e. vagueness as to where the permissibility-impermissibility boundary lies): there are multiple candidates for the reference of 'permissible' that are admissible by the lights of our normative principles and no candidate is the definite referent of 'permissible'. However, using Solution A we cannot explain why we don't know where the second order boundary (e.g. the boundary between definitely permissible acts and borderline permissible acts) lies. Explaining the vagueness of the second order boundaries would require postulating that there are multiple admissible sets of admissible candidates for any given conceptual role. However, then it seems that the union of all these sets would also form a set of admissible candidates – second-order vagueness would collapse into first-order vagueness. Thus, we must reject Solution A and place our hopes on Solution B.

6.2 Solution B

Fortunately, there is a way to give an epistemicist account of higher-order vagueness within this framework. Solution A requires adjusting the original principles proposed by Williams. On the other hand, Solution B requires slightly changing the notion of *conceptual role*, while leaving the rest of the framework intact. As so far described, the conceptual roles have been quite coarse-grained: many patterns of behaviour and inference have qualified as being the same conceptual role. Solution B relies on the idea that conceptual roles are taken to be more fine-grained: a narrower range of behavioural patterns counts as constituting the same conceptual role. The notion of a conceptual role is not a precise one: there is no precise boundary between communities that use the term 'wrong' in the same way that we do. On

Williams's account, 'wrongness' plays the *blame* assignment role in our community. However, it is vague what constitutes blame; for instance, it is unclear how strong a reactive response must be to constitute blame. Communities that differ slightly with the strength of the reactive response (e.g. in one community it is blame*, in another it is blame**), will differ slightly with regards to the conceptual role that 'wrongness' plays: in one community it will be the blame* assignment role and in the other the blame** assignment role.[57]

The heart of Solution B is this: small shifts in conceptual role across close possible worlds produce small shifts in the reference of moral terms that explain the vagueness in a standard way. On the fine-grained construal of conceptual roles, communities in nearby worlds will differ slightly with respect to the conceptual role that 'wrongness' plays. Conceptual Role Determinism holds for 'wrongness' in all nearby worlds – identity in conceptual role entails the same intension for 'wrongness'. Williams's normative principles dictate which property is denoted by 'wrongness' in each community: in a community where 'wrongness' plays the blame* assignment role 'wrongness' denotes wrongness*, in a community where 'wrongness' plays the blame** assignment role, 'wrongness' denotes wrongness** etc.

The solution to the initial puzzle then is this. Small shifts in the conceptual role of 'wrongness' across close possible worlds shift the reference of 'wrongness'. In a borderline case C, it is indefinite whether for some act A committed by x, A is wrong. The epistemicist explanation of vagueness is that the reason why C is a borderline case is that 'wrong' denotes different

[57] That is not to say that the intensity of the negative emotion is the only dimension alongside which we can soritise 'blame'. My goal is merely to show how the conceptual role in two communities can be very similar, but not identical (thus producing very similar, but distinct reference for moral terms).

properties in distinct communities. For instance, it denotes wrongness* in w* and wrongness** in w** that is close to w* and A-ing is wrong* but not wrong**. Suppose that blame* and blame** are reactive responses differing slightly in strength (blame* is slightly weaker than blame**). Suppose that in w* 'wrong' plays the blame* assignment conceptual role and in w** it plays the blame** assignment conceptual role. Then, a reason-responsive agent would be such that they would blame* x for A-ing but would not blame** x for A-ing (we can rephrase Williams's four normative principle using the notions of wrongness*/blame* and wrongness**/blame** respectively).

Solution B allows us to account for higher-order vagueness on this account. Small shifts in the conceptual role of 'wrong' shift the reference across the modal space. Once the shifts in the conceptual role of 'wrong' add up in some world v, it no longer counts as being relevantly close to the actual world. We are not in a position to know where the cut-off point between 'definitely wrong' and 'borderline wrong' is, because there are nearby (relevantly indistinguishable) worlds that v is close to (even though it is not close to the actual world). Thus, the same explanation (small shifts in conceptual role shifting the reference slightly) is applies to vagueness at different orders.

Now that we have the above metasemantics borrowed (and adapted) from Williams, we can try to go back to the initial objections posed by Schoenfield (2016) to explain why, despite their initial plausibility, the objections don't work. Schoenfield's objections presupposed a simple epistemicist model, on which it seems it would be possible to learn what moral facts are from the facts about usage. The epistemicist can object that we don't know what the usage facts are and how these facts alongside the metasemantic laws determine the meaning facts. However, that is not a reply that would convince critics like Schoenfield, who can insist that even though we don't in fact know the usage facts or the metasemantic laws, in principle that might be possible (especially

if we adopt an idealisation that it would be in principle knowable for an omniscient anthropologist or linguist).

Nevertheless, we can answer the objections more directly using the new metasemantics. On the metasemantics presented above, it is not possible to learn moral facts without engaging in moral deliberation (Williams 2018). It is not enough to know how the metasemantic rules determine the reference of moral expressions. Different moral theories will disagree over the role that 'wrong' plays in communities and over which acts are to be categorised as morally wrong. The disagreement over the first order moral theory will seep into the metasemantics of moral terms. The reference of moral terms is determined by which moral theory is correct, because only interpreting agents as referring to *wrongness* (when using the term 'wrong') would make them most reason-responsive. Determining which property that is, is the task of the first order moral theory.

Therefore, the Cheryl will not be able to learn after which N days abortion becomes impermissible after talking to an idealised anthropologist (unless the anthropologist has consulted some experts on morality who have established what the true moral theory is). Nevertheless, we have a way of accounting for vagueness as outlined above. Hence, by adopting the framework from Williams (2018) we have a semi-stable metasemantic theory for moral terms that is compatible with epistemicism while satisfying the demands presented by Schoenfield (2016) objections and the Moral Twin Earth cases. In the next section, I will consider potential problems for the account, and I will argue that these issues can be dealt with.

7 Potential problems

7.1 Fine-grained conceptual roles

Solution B relies on fine-grained conceptual roles. The difference between the two proposals should be framed as a difference between modally plastic and modally robust properties (Dorr & Hawthorne 2014). For instance, the relation $\lambda x \lambda y.$ *x is within the eyesight of y* is a modally robust relation: we could significantly change the distance between x and y and the relation would still hold (in many cases). On the other hand, the relation $\lambda x \lambda y.$ *x is exactly the same height as y* is not modally robust: changing the height of x even slightly means that it stops being exactly the same height as y. If the relation $\lambda x \lambda y.$ *x has the same conceptual role as y* is construed as a modally plastic relation, then we can easily give an account of moral vagueness using Solution B. On such a construal it is highly contingent what conceptual role a given moral term plays within a community. We can then use the original Williamsian normative principles and Conceptual Role Determinism thesis to explain vagueness of moral terms: slight shifts in conceptual role of moral terms slightly shift the reference of moral terms across linguistic communities. This gives us a uniform explanation of moral vagueness at every order.

One might worry that the construal of conceptual roles as fine-grained is unprincipled. But there is philosophical justification for this. Suppose that there is some far away possible world W* where the conceptual role of 'wrongness' is clearly distinct from the role it plays in our community. We assumed that the conceptual role of 'wrongness' is defined by its relationship to *blame*: all and only acts that are wrong deserve blame. When someone commits a wrong act, we blame them which means that we hold some negative feelings towards the perpetrator. In case of horrifically wrong acts, such as e.g.

starting an unjust war, our feelings towards the perpetrator are quite strong. Suppose that at W* the community's feelings towards perpetrators of horrific acts such as starting an unjust war are very mild, comparable to the negative feelings we might feel towards someone wearing unmatching socks. I think it is clear that the very slight negative emotion they feel towards horrific acts is not blame, it is something clearly distinct (say blame*). Their use of the term 'wrong' is associated with the link to blame* and not with blame – blame requires feelings of appropriate strength. So 'wrong' plays a very different conceptual role in the actual world and at W* and refers to a different property – not *wrongness* but some other property, e.g. *wrongness**. If this is the case, then there is a series of worlds between W* and the actual world, where the strength of the negative emotion towards wrong acts is slightly larger with every consecutive world. In such a case, it is natural to suppose that the conceptual role of 'wrong' in each world is slightly different, which produces small shifts in the reference of 'wrong' in each world. Of course, it is not inconsistent to suppose that conceptual roles are modally robust; however, such a construal is less natural in the current setting.

7.2 Controversial Metaethics

A potential worry for the epistemicist account outlined above is that it implicitly relies on a specific metaethical theory. In presentation of Solution B, I relied on the idea that communities in nearby worlds differ with respect to the conceptual role played by 'wrongness'. Crucially, I also relied on the idea that Williams's normative principles will select different properties as referents of 'wrongness' in communities where the conceptual role is slightly different, i.e. that the normative principle tells us that wrongness* is the appropriate referent for a community where 'wrongness' plays the blame* assignment role

and so on for other variants of blame. One might worry that this kind of picture might not be compatible with every kind of metaethical theory out there.

There are two things to be said about that worry. First, looking at the dialectical situation, the goal for the epistemicist is to resist the challenge to the theory as presented by Schoenfield (2016) that epistemicism is incompatible with moral vagueness. To meet the challenge, it is enough for the epistemicist to show that epistemicism is compatible with moral vagueness, even if the epistemicist solution relies on a particular metaethical theory. Epistemicism is compatible with moral vagueness. It may be the case that someone will formulate a challenge to epistemicism that would say that epistemicist treatment of moral vagueness is incompatible with some specific metaethical theory. However, so far, such a challenge has not been formulated, so there is no problem for the epistemicist to address.

Secondly, it is not clear that any challenge to epistemicism from metaethics will not just be a challenge to moral vagueness. For instance, some metaethicists might be unhappy about the fact that epistemicists predict that some moral truths are unknowable. However, many theories of vagueness predict that if ϕ is borderline, then ϕ is unknowable; such metaethicists should just deny that there is moral vagueness and claim that our ignorance in problematic cases (like Darryl's) is just a result of us not knowing enough about morality. Some metaethicists may be also unhappy about the fact that epistemicists predict that there are multiple properties that could potentially regulate the life of a community; however, again, the vast majority of theories of vagueness (including theories of ontic vagueness supported by Schoenfield) predict that for any vague term there are multiple eligible candidates for its reference (I expand on that point in the next section). So, there is no obvious challenge to the epistemicist that would not also be a challenge to moral vagueness in general.

7.3 Action-guidance (Schoenfield 2016: 273)

Another objection to the epistemicist account that I proposed is that my view cannot account for action-guiding features of moral properties. A version of this objection is posed by Schoenfield (2016: 273). The problem is the following. Epistemicism relies on the idea that a vague term like 'permissible' could easily have a different reference than it actually has. Suppose that 'permissible' actually denotes the property permissibility* but could easily denote some different but very similar property permissibility**. If there is a borderline case such that it is permissible* to ϕ but it is impermissible** to ϕ, it seems that there remains a question: 'what ought a rational agent do?'.

One thing to say here is that the question becomes much less puzzling if we attempt to precisify what is meant by 'ought'. If 'ought' is derived from 'permissible' so that one ought to ϕ if and only if it is impermissible not to ϕ, then both permissibility* and permissibility** will come with accompanying notions of ought* and ought**. The answer to the question will depend on the interpretation of 'ought': a rational agent ought** not to ϕ, but it is not the case that she ought* not to ϕ.

I think it is possible to press the objection further. After all, if there are multiple admissible interpretations of 'ought', there seems to remain a problem of how one ought to act in a borderline case, which is a specific instance of acting under indeterminacy.[58] However, this is not a particular problem for epistemicism, rather a more general problem for the proposition that there is moral vagueness. Moral vagueness commits us to the thesis that there are multiple equally fundamental candidates for the reference of 'permissible'. The ontic interpretation of moral vagueness proposed by Schoenfield (2016) does not fare any better here. For proponents of the ontic interpretation of moral

[58] See Williams (2014).

vagueness, it is indeterminate what one ought to do in a borderline case. If it is indeterminate whether one ought to ϕ (or as some put it 'there is no fact of the matter' as to what one ought to do) then we also have an action guiding problem: what ought one do if it is indeterminate whether one ought to ϕ? For the ontic vagueness theorist, the vague property *permissibility* has multiple eligible cut-off points; for the epistemicist, there are multiple interpretations of 'permissible' in different linguistic communities that denote different properties (permissibility*, permissibility** etc.). Prima facie, either theory will have to be supplemented with a theory of decision making and action under indeterminacy.[59] However, that would be an entirely different project; for now, it does not seem like epistemicism is faced with a particularly difficult task here.

The ontic vagueness theorist may attempt to underline the difference between the two theories. The ontic theory claims that there is only one moral property (in the vicinity), namely *permissibility*, whereas epistemicism claims that there are multiple such properties. However, the difference here between the two theories is immaterial. Both theories need the claim that there are multiple eligible candidates for the reference of permissible. Moreover, it cannot be a fundamental fact that one of the candidates is more eligible for reference (then the reference would be determinate). Where there is a difference between the ontic and the epistemic view is in communication. On the ontic view, different linguistic communities are assigned the same interpretation, so there are no issues with miscommunication resulting from the difference in interpretations. On the other hand, on the epistemic view different linguistic communities are assigned with different interpretations, which potentially results with miscommunication. However, that is a well-known feature of the

[59] See Williams (2017) for a survey of approaches to the problem.

epistemicist account and is by no means limited to the moral case. Therefore, moral vagueness does not present any special problems here.

Lastly, accepting that there are multiple eligible 'oughts' (or multiple eligible cut-offs for *permissibility*) may seem like a steep price to pay. However, it's a price to be paid by anyone who believes that there is moral vagueness: if ϕ is vague then there are multiple candidates for its reference. This includes non-epistemic theories. If one cannot bear to pay the price of multiple candidates for 'permissibility', one ought to reject moral vagueness.

8 Conclusion

In this chapter I presented a semi-stable metasemantic theory of moral terms. I made use of Robbie Williams's (2018) metasemantic framework to show that epistemicism is compatible with (a) moral vagueness and (b) Moral Twin Earth intuitions. I argued that there are two ways of combining Williams's framework with epistemicism and that one has an easier way with dealing with higher order vagueness. In the last section, I argued against potential challenges for the account. For all we know, moral vagueness just is epistemic vagueness.

Part II – Technical Issues

The second part of the thesis consists of three chapters and focuses on technical issues. My aim is to refine certain parts of the epistemicist theory. The Williamsonian epistemicism is often summed up in a slogan*: borderlineness is ignorance due to semantic plasticity*. How do we develop the theory based on our slogan? The recent literature on epistemicism (Caie 2012, Magidor 2018, Yli-Vakkuri 2016) has proceeded by developing systematic accounts of the definiteness operator Δ that would express the epistemicist treatment of definiteness. However, it seems to me that a better starting point for refining the slogan would be to answer the question: what does it mean for ignorance to be due to a particular factor (like semantic plasticity)? The literature so far has relied on an intuitive understanding of factor-relative ignorance. The epistemicist explains ignorance by failures of safety. In Chapter 4, I try to systematise the notion of safety relative to a factor. It turns out that we can provide a basic account of factor-relative safety; however, the account would need significant refinement for us to be able to straightforwardly derive the semantics for the epistemicist definiteness operator.

Chapter 5 develops an account of the definiteness operator for epistemicism inspired (though not directly derived from) the account of factor-relative safety developed in Chapter 4. The account uses a multidimensional model, similar

to the account developed by Litland & Yli-Vakkuri (2016), but improves on the work that exists in the literature as it avoids certain important problems that have hindered the development of the epistemicist definiteness operator so far.

Chapter 6 closes of the technical part of the thesis. It turns out that sentences involving metalinguistic vocabulary such as 'truth' pose an important challenge to presenting an account of the definiteness operator. Chapter 5 only sketches a treatment of the semantics of 'truth' that are required by the account of definiteness. In Chapter 6 I go into greater detail. I use a puzzle due to John Hawthorne (2006) to show what restrictions are imposed on the notion of truth by the epistemicist theory (including the account of definiteness developed in Chapter 5).

Chapter 4

Factor-relative Safety

ABSTRACT

The epistemicist theory of vagueness explains ignorance in borderline cases by failures of safety. More specifically, the epistemicist argues that our beliefs in borderline cases are not safe relative to the variation in intensions assigned to expressions in our language across nearby possible worlds. My aim in this chapter is to provide a broader framework for analysing factor-relative safety.

WORD COUNT: 10,012

1 Introduction

According to epistemicism, vagueness gives rise to a particular kind of ignorance. The basic story says that the semantic plasticity of vague expressions and concepts we use is an obstacle to knowledge. In a slogan: ignorance due to vagueness is ignorance due to semantic plasticity.

The epistemicist claim that ignorance in borderline cases is due to semantic plasticity raises a natural question: what is it for someone's ignorance to be *due to* a particular factor? We can be ignorant of a particular fact for many reasons. For instance, we may be ignorant whether Michael is tall, because Michael is a borderline case of tallness; alternatively, we may be ignorant

whether Michael is tall, because we only see Michael from a distance and cannot exactly assess his height. Vagueness is responsible for our ignorance in the former case, but not in the latter.

In this chapter I will attempt to provide a general framework for how to understand the claims that one's ignorance is due to a particular factor.

2 The safety condition on knowledge

2.1 Simple safety

The Williamsonian epistemicist relies on a safety-based account of knowledge. It's often considered an advantage of the epistemicist view that it can predict the existence of borderline cases using a more general (and independently motivated) safety-based framework. Therefore, it's important to spell out what the epistemicist is committed to. The general idea of safety-based account of knowledge is that knowledge requires safety from easy error. The rough idea can be stated as follows:

(Safety) If X knows that p, then X could not be easily wrong (about p).

Suppose that X knows that p. If X knows that p, then X believes that p. Additionally, Safety requires that X's belief be safe: there is no close possible world where X's belief is false.[60]

[60] One might wonder what counts as a close possible world in this context. One requirement that is usually made is that worlds count as close only if they are epistemically possible. A world w is epistemically possible if and only if for all we know we are in w. The reason why close worlds have to be epistemically possible is simple. To see whether X could easily make an error, i.e. easily have a false belief about p, we should only look at worlds where X has that belief (if a world is not epistemically possible there is no guarantee X has that belief in that world). We could also model safety so that there are nearby worlds that are not epistemically possible; then if X knows p, it does not mean that p is true in every nearby world, but true in all nearby worlds in which X has the belief. The latter option is more

On a simple model, to see whether X's belief that p is safe we should just determine whether p is true in every close possible world. On that model, safety behaves as a kind of restricted necessity (Williamson 2009, p. 14). However, there is a problem with such an approach if we are to use it for the epistemicist's needs. Firstly, the content of a belief may vary between close worlds. Suppose that I express my actual belief using the sentence 'Michael is tall'. In the actual world that sentence picks out some proposition p, which says that *Michael is tall*. However, in a close world w^* where the language is used slightly differently, the sentence 'Michael is tall' expresses some different proposition p^*, which says that *Michael is tall**. Since I would be disposed to express my belief at w^* using the sentence 'Michael is tall' just as I am disposed to do so in the actual world, the relevant proposition to evaluate at w^* is not p (the proposition I actually believe) but rather p^* (the proposition I would believe at w^*). Expressing my belief at w^* using the same sentence as I do actually runs the risk that I will say something false, because of the shift in the proposition expressed by the sentence.[61] So instead of evaluating the same

difficult to model, so more often then not the former option is preferred in modeling safety.

[61] Considering what proposition I would have believed in w^* as opposed to taking the proposition I actually believe and evaluating it at w^* is independently motivated. Suppose that I am using a measuring stick to measure the length of a log. It turns out that the company producing the logs is very precise and there is virtually no variation in the length of the logs they produce. If I measure the log to be 10 meters, it could not easily be the case that the length of the log be different. However, suppose that the company that produces measuring sticks is not as precise and that consequently the measuring stick that I have could easily be of a different length (the length of 1 meter on the actual stick is different from the length of its counterpart in a nearby world). Intuitively, my belief that the log is 10 meters in length is not safe (even if we assume that my actual stick is by accident produced correctly so that '1 meter' on my measuring stick corresponds to the length of 1 meter). The belief is not safe even if there is no variation across nearby worlds with regards to the length of the log, because the variation in the length of the measuring stick could easily produce an error. If my measuring stick could easily be of different length, there is a nearby world where (using my measuring stick) I judge the log to be 11 meters in

proposition at close possible worlds, we have to complicate the picture slightly and instead evaluate propositions that would be believed had those worlds been actual. For this reason, namely that the content of our beliefs may be slightly different in nearby worlds, the epistemicist is often interpreted with being committed to the following Metasemantic Safety Principle:

> (Metasemantic Safety) If X knows that p under the guise ϕ then there is no close possible world w such that ϕ as used in w is false in w.[62]

The Metasemantic Safety principle uses the connection between beliefs and sentences that express them: it presupposes that the contents of sentences at close possible worlds match the contents of beliefs an agent would have at those worlds. In other words, the contents of sentences at close worlds serve as a proxy for contents of beliefs one would have at those worlds.[63]

---

length and that judgement is wrong. Consequently, in order to show that my belief is unsafe, we have to look at the truth value of the proposition I would have believed in a close possible world (and not merely at the proposition that I actually believe).

[62] It may seem that replacing Safety with Metasemantic Safety is an ad hoc fix required by the epistemicist, but there are independent reasons for Metasemantic Safety. For instance, suppose that we introduce into the language the Kaplanian (1978) operator *dthat*, which rigidifies descriptions. Suppose that I form the belief that the ticket dthat will win the lottery will win the lottery. It seems that I am in a position to know that the ticket dthat will win the lottery will win the lottery. But to evaluate my belief as safe, at every close world we have to evaluate the proposition that I would believe at that world since in every world the ticket that wins the lottery changes.

[63] This is important to note because some of the criticisms of the Metasemantic Safety principle seem to miss this point. For instance Bacon (2018) provides the following putative counterexample to Metasemantic Safety. As a linguistic community we could easily decide to use our numerical symbols differently, e.g. we could easily have chosen to use the symbol '1' to refer to the number 100. But it does not mean that we are not in a position to know that 1+1=2. This is not a problem for the Metasemantic Safety principle. If in some world w we use '1' to refer to 100, I would not be disposed to express my belief that 1+1=2 using the expression '1+1=2'. Therefore, such a world would not count as a close world since it is not epistemically possible (even though there is an intuitive sense in which we could easily use the expressions in our language very differently).

2.2 The multi-factor problem

The second worry about the simple model is that when safety is violated, so far there is no way to say why it was violated. The epistemicist associates ignorance in borderline cases with semantic plasticity. If we had used our language slightly differently, the vague expressions in our language would have slightly different boundaries. Semantic plasticity imposes a particular kind of risk of error on our beliefs: even if my belief in a borderline case is actually true, the belief I would have in some close world would be false. Suppose that Michael is a borderline case of 'tallness'. Even if my belief expressed by the sentence 'Michael is tall' is actually true, in some close world I would still be disposed to express my belief by the sentence 'Michael is tall', but it would express some different and false proposition at that world. Therefore, the epistemicist associates the ignorance in borderline cases with a particular ignorance-inducing factor: semantic plasticity. So far, the simple model on which we evaluate the beliefs (or their counterparts) at close possible worlds is only able to tell us whether they are true or false at those worlds, but unable to tell us what is responsible for our ignorance.

To illustrate the issue consider the following case[64]:

> (Definiteness Puzzle) Suppose that Michael is 190 cm in height. Moreover, suppose that the actual cut-off point for 'tall' is 185 cm: anyone over that height counts as 'tall'. Furthermore, suppose that the relevant margin of error associated with the semantic plasticity of 'tall' is 3 cm. Consequently, only the linguistic communities that place the cut-off point that differs from the actual cut-off by 3 cm count as relevantly close (relative to the actual world). Suppose that there is a nearby world $w*$ where the cut-off point for 'tall' is at 188 cm, but

[64] The puzzle is due to Ofra Magidor (2018).

where Michael is 187 cm in height (only slightly shorter than he actually is). Then Michael in w^* does not count as being 'tall' (as used at w^*).

This puzzle presents a challenge to epistemicism: Michael was supposed to be a definite case of 'tallness' and yet the case shows that 'Michael is tall' is not metasemantically safe. Since there is a close world where 'Michael is tall' is false, then by Metasemantic Safety we don't know that Michael is tall. Intuitively, in the above case the semantic plasticity of 'tall' is only partly responsible for our ignorance. The shift in the intension of 'tall' is not sufficient by itself to make Michael fall out of the extension of 'tall', so the responsibility for our ignorance cannot be placed entirely on semantic plasticity. In the above case, the assumption is that w^*, a world that differs from the actual world with respect to Michael's height, counts as a nearby world. The change in the underlying non-semantic facts (Michael's height) also contributes to Michael falling out of the extension of 'tall' at w^*. The reason why this is possible is that semantic plasticity is not the only ignorance-inducing factor there is. There are other factors at play. For instance, our judgment regarding assessing a person's height in cm may be inexact, because we are not using a measuring tape but only our eyesight. If we had exact knowledge of Michael's height in cm, w^* would not count as a nearby world and we would be in a position to know that Michael is tall.

Whether this is a problem for epistemicism depends on the version of epistemicism we adopt. In the next section I will consider two versions of epistemicism that will have different approaches to the multi-factor problem.

2.3 Ambitious and Non-ambitious Epistemicism

The severity of the multi-factor problem for epistemicism depends on how tight the relationship between semantic plasticity and ignorance is needed for the

epistemicist purposes. This in turn depends on the version of epistemicism we adopt. The multi-factor problem poses an issue for the following version of epistemicism:

> (Ambitious Epistemicism) Borderline cases give rise to ignorance. This ignorance in is explained by failures of safety *due to semantic plasticity*.

The simple model cannot tell us whether a failure of safety is due to a particular factor: it cannot tell us whether safety fails due to semantic plasticity or due to some other factor. However, this may not be an issue for a slightly different version of epistemicism:

> (Non-Ambitious Epistemicism) Borderline cases give rise to ignorance. This ignorance in is explained by failures of safety.

How to understand these two positions?

The Ambitious and Non-ambitious versions of epistemicism disagree how seriously to treat the notion of ignorance due to semantic plasticity. Ambitious Epistemicism treats it seriously: the epistemicist has to explain what it means for our ignorance to be due to a particular factor like semantic plasticity. On the other hand, Non-Ambitious Epistemicism does not treat that challenge seriously: what the epistemicist has to explain is ignorance (not ignorance due to a particular factor).

The difference between the two positions is subtle. The aim of Non-ambitious Epistemicism is to present a minimal version of the theory that would explain all that is required from a theory of vagueness, but would remain quiet about all other issues. From the Non-ambitious perspective the epistemicist does not have to explain very much when it comes to vagueness-related phenomena. The epistemicist keeps classical logic as well as semantics, which don't call for much explanation. Moreover, it presents a straightforward

answer to the Sorites paradox: for any vague expression there is a sharp boundary, it's just that we are not in a position to know what it is.[65] What is left for the epistemicist to explain (a) why we are not in a position to know where these sharp boundaries lie and (b) why we are ignorant in borderline cases. The standard epistemicist reply answers both (a) and (b): semantic plasticity prevents us from having exact knowledge of the sharp boundaries of vague terms and prevents us from knowing in borderline cases. However, both the challenge and the reply may be formulated in slightly different ways.

One way of reading the challenge to epistemicism is to just explain the ignorance. Vagueness gives rise to ignorance. What is characteristic about vagueness is that there is an ignorance-inducing factor, semantic plasticity, that plays part in the explanation of ignorance in borderline cases. But the ignorance in particular borderline cases may be explained using a multitude of factors. If there are more ignorance-inducing factors (like in the case of estimating Michael's height), then ignorance is less puzzling and easier to explain. If our goal is only to explain ignorance (and not ignorance due to a particular factor like semantic plasticity), then Definiteness Puzzle does not present any challenges. In the case outlined in the Definiteness Puzzle, there are at least two factors that can produce failures of safety: the variation in meanings of expressions (semantic plasticity) and the variation in height. If all that we have to explain is ignorance, then the task is easier if there are more factors that we can appeal to. Therefore, it is not as if Non-Ambitious Epistemicism does not claim that semantic plasticity plays a role in the explanation of ignorance. It is just that the Non-Ambitious Epistemicist does

[65] One might object that this does not yet constitute a full solution to the Sorites; a full solution requires explaining why the Sorites premises are compelling in the first place. I believe that the explanation for why the Sorites premises are compelling is a different task and one that is likely to be independent of the particular view of vagueness one holds.

not think it's their task to explain what it means for ignorance to be produced by a particular factor.

On the other hand, on the 'ambitious' reading of the challenge to epistemicism, the epistemicist must not only explain the ignorance (of boundaries and in borderline cases), but also analyse the notion of safety-relative to semantic plasticity. It is one thing to demand from the epistemicist to explain how safety is violated and another to explain how safety is violated solely due to semantic plasticity.

One way of cashing out this difference between Non-Ambitious and Ambitious Epistemicism is by focusing on how seriously they treat the borderline/definite distinction. For the Non-Ambitious Epistemicist, the distinction is not an important, joint-carving category. In a slogan, the Non-Ambitious Epistemicist adopts an ignorance-first approach. What is important is the distinction between knowledge and ignorance, not the distinction between a specific kind of ignorance (ignorance due to semantic plasticity) and safety with regards to a particular factor (definiteness). Once the Non-Ambitious Epistemicist is able to explain the ignorance in borderline cases and the ignorance of sharp boundaries, that job is done.

On the other hand, the Ambitious Epistemicist adopts an indefiniteness-first approach. What is required of a theory of vagueness such as epistemicism is to explain what it means for a sentence (or a proposition) to be borderline or definite. The distinction between knowledge and ignorance is important (borderlineness gives rise to ignorance after all), but the crucial task for epistemicism is to explain borderlineness and definiteness. Definiteness is safety from error due to semantic plasticity. Thus, the task for the Ambitious Epistemicist is to model ignorance due to semantic plasticity.

The names that I have given to the two versions of the epistemicist theory may seem tendentious: who would want to subscribe to a non-ambitious

theory? However, I do believe that the Non-Ambitious stance is a genuine option for the epistemicist. Unlike most other theories of vagueness, epistemicism does not have to treat borderlineness and definiteness as *sui generis* phenomena. The epistemicist simply notes that there is a feature of some expressions in our language, semantic plasticity, and shows that there are instances where this feature will give rise to ignorance. From the epistemicist perspective, it's perfectly reasonable to insist that definiteness and borderlineness are superfluous categories.

The distinction between Ambitious and Non-ambitious Epistemicism is important, because depending on our choice of the theory, we can treat the multi-factor problem, illustrated by the Definiteness Puzzle, differently. The multi-factor problem is an issue for the Ambitious Epistemicist. According to Ambitious Epistemicism, semantic plasticity is responsible for ignorance in borderline cases and ignorance of the sharp boundaries. This puts pressure on the Ambitious Epistemicist to explain what it means for ignorance to be due to a particular factor like semantic plasticity. The difficulty is to show how to delineate the effect of a particular ignorance-inducing factor. The Definiteness Puzzle is an illustration of this difficulty: sometimes safety is violated because of a combination of factors.

On the other hand, the Non-Ambitious Epistemicist does not seem to have a problem with the Definiteness Puzzle. The epistemicist task, according to Non-Ambitious Epistemicism, is simply to explain our ignorance, not necessarily to attribute that ignorance to a particular factor. The Non-Ambitious Epistemicist will invoke semantic plasticity to explain our ignorance in particularly problematic cases. For instance, semantic plasticity as an ignorance-inducing factor is very useful in explaining why we don't know whether Michael is tall, when he is a borderline case of 'tall', even though we know his exact height in cm. However, it's much easier to explain why we don't

know whether Michael is tall if we don't know his exact height in cm. It's easier to explain why we are ignorant if there are more ignorance-inducing factors at play.

My strategy for this chapter is to pursue the Ambitious Epistemicist line. There are dialectical reasons to prefer Ambitious Epistemicism to its non-ambitious cousin. Since other theories of vagueness treat borderlineness and definiteness as theoretically joint-carving phenomena, it's useful for the epistemicist to be able to paraphrase the talk about borderlineness and definiteness in epistemicist terms just to remain a part of the conversation (instead of inviting even more incredulous stares). Therefore, I will pursue assuming that modelling ignorance-due-to-semantic-plasticity is a worthwhile task for the epistemicist.

3 Factor-relative safety

3.1 The challenge

The Definiteness Puzzle outlined above presents a situation where a combination of multiple ignorance-inducing factors produces ignorance. Semantic plasticity is only partly responsible for the ignorance, because the variation in intensions assigned to expressions across the space of close possible worlds is insufficient by itself to produce ignorance. Michael is 190 cm in height, the boundary for 'tall' is 185 cm and the margin of error for semantic plasticity is 3 cm. Clearly, the variation in the intension assigned to 'tall' alone is insufficient to shift the truth value of 'Michael is tall' from true to false. Thus, we may say that a belief that Michael is tall is safe relative to semantic plasticity: variation in this ignorance-inducing factor alone is insufficient to produce an error. We can generalise this observation as follows:

> (Factor-relative Safety) ϕ is safe relative to F if and only if the variation in F alone is insufficient to easily render ϕ false.

How to represent this factor-relative safety within the model developed thus far?

What we would ideally like, while keeping everything else constant, is to shift the value of our factor of interest to see whether shifting the factor alone would be sufficient to render ϕ false. For instance, in case of the Definiteness Puzzle, to see whether 'Michael is tall' is safe relative to semantic plasticity we would want to look at situations where only the intension of 'Michael is tall' shifts while everything else (like Michael's height) is kept constant. One way to do this is to try to isolate the effects of a single factor by looking at nearby worlds that differ from the actual world only with regards to the value of the factor in question.

In the next section, I will present a simple model that allows us to do it. The development of the simple model will allow us to see the notion of factor-relative safety in operation and formulate some interesting observations about it. Furthermore, we will see the limitations of the model, which will allow us to develop a more complex model free of the limitations.

4 The simple model

My analysis is meant to be applied to the epistemic notion of safety, but it can also be extended to non-epistemic notions of safety. Let's consider a simple example that will help us in the analysis. Suppose that I took a test and got a score of 80. My test-scoring ability depends on various factors; for instance, it depends on how well-rested I am on the day of the test. Suppose that the factor *how well-rested I am on the day of the test* can easily change my score by up to 5 points. This means that relative to how well-rested I am on the day of the

test, my score is safely above 75. This does not mean that overall my score is safely above 75; for instance, if it could easily be the case that the test is more difficult, then my score could easily vary from my actual score by more than 5 points. But the difficulty of the test is a different factor.

Suppose that we want to check whether p is safe at w relative to some factor F. To check that we would like to see what value F takes in all the nearby worlds and whether it can make p false. However, we would like to keep the facts that are not relevant to F fixed, e.g. if p is a proposition completely unrelated to F, then we would like p's truth value not to change as a result of changing F (even if p's truth value changes in nearby world as a result of a change in some other factor).

The idea is, for the purposes of our analysis, to carve out two parts from possible worlds: one that corresponds to the F-aspects of the world and one that corresponds to the aspects of the world unrelated to F. Suppose that we think of possible worlds as maximally compossible sets of propositions. Then we can carve out two sets of propositions out of w: F(w) is the set of propositions that describe the F-aspects of w and F'(w) is the 'anchor' set of propositions unrelated to F. For instance, when we are interested in analysing whether the score of my test is safely over 75 relative to the variation in how tired I am on the day of the test, we will carve up the world w so that F(w) is the set of propositions describing how tired I am on the day of the test at w. To check whether the proposition that my score is over 75 is safe at w relative to F, we would fix (anchor) all the facts that are not relevant to F and allow only for the F-relevant facts to shift. We would evaluate the proposition only in nearby worlds that satisfy all the propositions in F'(w) (the set of propositions true in w that are not concerned with F). Similar procedure would allow to spell out safety relative to other potential ignorance inducing factors.

A natural starting point in our attempts to model factor-relative safety is to explain how to conceive the factors relative to which propositions might be safe. We can conceive of our factors as subject matters (in the sense of Lewis 1988a, 1988b) that partition the logical space into equivalence classes. If we think of subject matters as questions, for instance 'what is the value of F?', then this subject matter will partition the worlds into equivalence classes (cells): each world from a cell will have the same answer to the question as all the other members of the cell. For instance, if F is the subject matter 'how tired am I on the day of the test?', then w and v are in the same cell under partition F if and only if on the day of the test I am as tired in w as in v.

Consequently, we ought to explain how to divide a possible world using a factor F into the desired two parts (a part related to F and a part unrelated to F). Let F(w) be the set of propositions that are true in all the members of the cell that w is a member of under partition F. Intuitively F(w) includes the minimal proposition that is true in all the worlds from the cell (call it the minimal proposition of F(w)), which we can interpret as the answer to the question posed by the subject matter, and its logical consequences. Let F'(w) be our anchor: intuitively the part of the world that is not about the subject matter F. The construction of F'(w) is a bit tricky.

Let's start with F'(w) including all the propositions that are true in w and proceed by removing the ones that should not be there. Firstly, all the propositions that are in F(w) (propositions wholly about F and their logical consequences) should be removed from F'(w). Secondly, all the propositions that are partly about F should be removed from F'(w).[66] Suppose that:

[66] On the Lewisian account of aboutness, a proposition p is wholly about a subject matter F, if and only if p always has the same truth value in F-equivalent worlds, i.e. for any cell, p has the same truth value in all the worlds in a given cell. I am using the notion of *being partly about a subject matter* in the following sense: p is

- p is the proposition *I am 10% tired*
- q is the proposition *my car is red*

If the subject matter is 'how tired I am' and $p \wedge q$ is true in w, we want to remove it from F'(w). Even though this proposition is not true in all the worlds in w's cell, because in some worlds in the cell my car is not red, and thus is not a member of F(w), we have to remove it from F'(w) as the variation in the subject matter (Michael's height) affects the truth of this proposition.

One important consequence of this construction is that if we make the subject matter F overly specific, F'(w) will be empty. Suppose that under partition by F, w is the only member of its cell. Then F(w) includes all the propositions that are true in w and F'(w) is empty. This turns out to be a limiting case that will correspond to safety *simpliciter* (safety relative to any factor). On the other hand, if the subject matter F is so general (or degenerate in the parlance of Lewis (1988b)) that all possible worlds are members of the same cell, then F'(w) will include every contingent proposition that is true in w. In such a case only, necessarily true propositions will be members of F(w). Since there is only one cell under the partition by a degenerate subject matter, there are no contingent propositions p such that either p or its negation is true in every world in the cell. Therefore, F'(w) includes all the propositions contingently true in w. Such a partition corresponds to safety relative to no factor: what happens if we keep every factor fixed.

Let's outline a simple model that would show the basics of the proposal. I will use the standard possible worlds account of propositions in this model. Let our model be a quadruple M = $\langle$W, R, $\mathcal{F}$, V$\rangle$, where:

- W as usual is a set of possible worlds,

---

partly about subject matter F, if and only if there is a cell under the partition by F such that p has the same truth value for all the worlds in that cell.

- R is a reflexive accessibility relation,
- $\mathcal{F}$ is the set of factors: F $\in \mathcal{F}$ iff F is a partition of W,
- V is a function from worlds to sets of propositions.

Members of $\mathcal{F}$ are subject matters that partition the logical space W into equivalence classes (cells), i.e. each member of $\mathcal{F}$ divides the logical space into sets of worlds that are disjoint and jointly exhaustive of W. Any two worlds w and v are in the same cell under partition F if they are exactly alike when it comes to the subject matter F. For instance, if F is the subject matter *variant of the language*, then w and v are in the same cell under partition F if and only if the variant of the language (the interpretation assigning intensions to expressions) is identical in w and v.

Let $c_F(w)$ be the cell (a set of worlds) that w is a member of under partition by F and let $F(w)$ be the set of propositions that are true in all the worlds that are members of $c_F(w)$.[67] Let F'(w) be the anchor, constructed as discussed above. The interpretation function V gives us, for each world w, a set of propositions true at w.

The bivalent valuation is defined as follows:

$|p|_w = 1$ iff $p \in V(w)$

Logical consequence is defined as truth preservation in all worlds in all models. To complete our basic model of factor relative safety, we need a way of specifying a factor relative safety condition. We can do so by introducing an operator into the language that would represent factor relative safety. Let S be such an operator. If for some p, $S(p, F)$ is true then p is safe relative to some factor F. We can define S as follows:

$|S(p, F)|_w = 1$ iff for all $v \in W$ such that vRw :

(i) $v \in c_F(w)$ and $|p|_w = 1$; OR

[67] We can define $F(w)$ as $\{p: \forall v \in c_F(w),\ p \in V(v)\}$.

(ii) if for all $q \in$ F'(w), $q \in$ V(v) then $p \in$ V(v)

The definition of S works as follows.[68] Firstly, p is safe at w relative to factor F, if F is such a general subject matter that all the nearby worlds belong to the same cell as w (e.g. if F is the subject matter regarding necessary truths such as which mathematical theorems hold). This simply means that p is safe because a change in the value of F could not easily occur: there is no nearby world where such a change occurs, so naturally such a change could not produce a failure of safety. Secondly, p is safe at w relative to factor F, if for any nearby world v that satisfies all the propositions that belong to the anchor F'(w), p is true at v.

Here is the picture. We would like to be able to analyse claims like 'p is safe at w relative to F'. The way to model this is by keeping the facts that are not related to F fixed and looking at the value of F in nearby possible worlds v. The goal is to see whether a variation in our factor F alone is sufficient to make p false. This simple model I outlined allows us to express the basic notion of factor-relative safety.

We can add the standard compositional truth conditions for conjunction and negation to allow for construction of more complex sentences:

$|\neg\phi\ |_w = 1$ iff $|\phi|_w = 0$

$|\phi \wedge \psi|_w = 1$ iff $|\phi|_w = 1$ and $|\psi|_w = 1$

Furthermore, there are several observations about the defined notion of factor relative safety that we can make.

OBSERVATION 1. S is factive: if $S(p, \text{F})$ is true then p is true.[69]

---

[68] As defined above, S is not iterable as we require ϕ not to include S. However, we can make S iterable as follows. Let S^n denote n iterations of S and let R^n correspond to the relation between worlds that are accessible in at most n steps (e.g. $w\text{R}^2v$ iff $w\text{R}v$ or there is a $u \in$ W such that $v\text{R}u$ and $u\text{R}w$). By replacing S with S^n and R with R^n in the above definition of S we get an iterable operator.

[69] This follows from the reflexivity of our accessibility relation.

OBSERVATION 2. S obeys Distribution: if $S(p, \mathrm{F})$ and $S(p \rightarrow q, \mathrm{F})$, then $S(q, \mathrm{F})$.[70]

OBSERVATION 3. If the subject matter F includes (in the sense of Lewis 1988b)[71] the subject matter G, then if $S(p, \mathrm{F})$ then $S(p, \mathrm{G})$.[72] Suppose that we are trying to assess whether my score on a test is safely above 80 relative to the variation in two factors: how difficult the test is (call it G) and how tired I am on the day of the test (call it H). Let F be the subject matter *how tired am I?* and *how difficult is the test?* If it turns out that p is safe relative to F, then p is safe relative to both G and H, i.e. if p is safe when we allow for variation in both of these factors, then it's safe when we consider the variation in only one factor.

OBSERVATION 4. $S(p, \mathrm{F})$ and $S(p, \mathrm{G})$ don't jointly entail $S(p, \mathrm{F\&G})$.[73] Suppose that we are trying to determine whether my test score is safely

---

[70] Proof. Case 1: if F is general enough that all nearby worlds belong to the same cell, the proof is trivial: if p and $p \rightarrow q$ are true in w, then q is too. Case 2: if for all nearby worlds v that satisfy all the propositions from F'(w), $p \in \mathrm{V}(v)$ and $p \rightarrow q \in \mathrm{V}(v)$, then $q \in \mathrm{V}(v)$ since possible worlds are logically consistent.

[71] The subject matter F includes the subject matter G if and only if for any two worlds, if they are equivalent relative to F than they are equivalent relative to G. Lewis's example is the subject matter *17th Century* and *The 1680s*: if any pair of worlds is identical with respect to events that occurred in the 17th century, then they are identical with respect to what happened in the 1680s.

[72] Proof. All we have to do is to prove that G'(w) contains all the propositions that F'(w) contains: if p is in F'(w) then p is in G'(w). We can prove its contraposition. Suppose that that p is not in G'(w). Case 1: p is in G(w). If p is in G(w), then p is in F(w), because the minimal proposition of F(w) entails all the propositions that are in G(w). Consequently, p is not in F'(w), because it is in F(w). Case 2: p is not in G(w), but it is not in G'(w), because it is partly about G: there is a cell under the partition by G that entails either p or its negation. However, if there is such a G-cell then there is an F-cell, that entails p or its negation (since F includes G). Therefore, p is not in F'(w). In either case, if p is not in G'(w), then it is not in F'(w).

[73] Proof. Let F and G be two orthogonal factors and F&G, be the factor that includes both F and G. Suppose that $p = q \vee r$ such that q is wholly about F and r wholly about G. Furthermore, suppose that both r and q are true in w, but there is a nearby world w^* where both r and q are false. For any nearby world v, $q \vee r$ will be both in F'(w) as well as in G'(w). Since r is wholly about G (which is orthogonal to F), r is in F'(w). By mirror reasoning, $q \vee r$ belongs to G'(w). However, under the partition by F&G, neither q nor r will remain in F&G'(w) and we know that there is a nearby world w^* where neither q nor r are true. Therefore, safety relative to F and safety relative to G, don't jointly entail safety relative to F&G.

over 75. My score is 80. There are two relevant factors: how tired I am on the day of the test and the difficulty of the test. Each of these factors has a margin of error of 3 points, i.e. across the nearby worlds the difference in my tiredness can shift my score by up to 3 points and the variation in the difficulty of the test can increase or decrease my score by up to 3 points. My score is such that it counts as safely over 75 with regards to each of the factors: my score is 5 points over 75 and either of the factors can produce a shift of at most 3 points. However, when the factors are considered in conjunction, they can produce an effect of up to 6 points: this would take my score below the 75 points threshold.
OBSERVATION 5. For any p if for any subject matter F, $S(p, \mathrm{F})$, then p is safe *simpliciter*.[74]

The model also has some important limitations. One issue is that we may have some problem applying the model to the initial case that motivated us to develop this account in the first place: safety-relative to semantic plasticity. In the case of semantic plasticity employing our basic model would mean that the worlds that count as close would be those that differ only with regards to the assignment of intensions to expressions in our language. However, semantic facts (facts about the assignment of intensions to expressions in the language at each world) presumably depend on non-semantic facts.[75] So we cannot keep all the non-semantic facts fixed, because this would mean keeping all the

[74] The maximal subject matter M, i.e. subject matter that includes all the others, would partition the logical space into singleton-sized cells. On such a partition M'(w) is empty and M(w) contains all the propositions that are true at w. This corresponds to the traditional notion of safety where we just evaluate propositions relative to every nearby world without holding any facts fixed.

[75] For a dissenting view see Kearns & Magidor (2012). My approach in modelling factor-relative safety will be ecumenical – ideally we would like a model which would be compatible with the thesis that semantic facts supervene on the non-semantic as well as it's negation. Nevertheless, the problem is quite general: some factor may be an obstacle to knowledge in case of the supervenient facts, even though it is not an obstacle to knowledge of the facts forming the supervenience base.

semantic facts fixed. Therefore, we cannot restrict the relevant set of worlds so easily.

The simple model uses a coarse-grained account of propositions which means that it treats necessarily equivalent propositions as identical. This is a problem because in some cases two necessarily equivalent propositions may behave differently when it comes to their epistemic status. Suppose that p is the proposition that *'tallness' refers to the property of being over 180 cm in height* and q is a proposition expressing some precise description of the relevant usage facts such that $\Box(p \leftrightarrow q)$. Even though p and q are necessarily equivalent, it seems that a factor like semantic plasticity should treat them differently: semantic plasticity seems to be an obstacle in coming to know p, but not q (since q is expressed by a precise description of the relevant usage facts). Ideally what we would like, when analysing safety relative to semantic plasticity, is to keep q fixed; however, if we keep q fixed, we also keep p fixed since they are necessarily equivalent. Therefore, to deal with these more complex cases we need a more complex model that would allow us to vary propositions like p while keeping propositions like q fixed. This will be my aim in the next section.

5 The complex model

On the simple model, we evaluated propositions at possible worlds. On the complex model we will do things slightly differently. Suppose that we were to evaluate sentences at interpretation-world pairs; then, we would be able to shift the values of one factor (e.g. the variance in interpretation due to semantic plasticity) and keep the other facts fixed. Something like this approach is employed by John Hawthorne (2006, p. 188) in his analysis of epistemicism. Instead of evaluating sentences at all nearby possible worlds, Hawthorne imagines that there are speakers using non-actual variants of the language in

the actual world. Thus, he is able to keep all the non-semantic facts about the actual world fixed and vary only the semantic facts. The world imagined by Hawthorne where all the non-semantic facts are kept fixed, but the semantic facts vary is not in fact a metaphysically possible world, if we assume that the semantic supervenes on the non-semantic. But the combination of actual non-semantic facts and non-actual semantic facts is what we need to evaluate safety-relative-to-semantic-plasticity claims.

We can attempt to model factor-relative safety on Hawthorne's (2006) approach to analysing semantic plasticity.[76] Hawthorne seems to take a 'part' of a nearby world (a variant of the language spoken at that world) and combines it with the anchor from the actual world (the actual world with the 'part' corresponding to the variant of the language removed). My approach will be to generalise this strategy. Suppose that we want to check whether p is safe at w relative to some factor F. To check that we would like to see what value F takes in all the nearby worlds and whether it can make p false. However, we would like to keep the facts that are not relevant to F fixed, e.g. if p is a proposition completely unrelated to F, then we would like p's truth value not to change as a result of changing F (even if p's truth value changes in nearby world as a result of a change in some other factor).

The idea is again to carve up the world into two parts using a factor: F(w) and the anchor F'(w). On the basic model, we evaluated propositions at nearby possible worlds that satisfied all the propositions in F'(w). Here we are going to do something different. We will evaluate the proposition relative to a set of propositions constructed from parts of possible worlds: a part that corresponds to the value of the factor F that we vary as we go across the set of nearby

[76] I don't claim that Hawthorne (2006) committed himself to impossible worlds; nevertheless, the way he sets up his Persons' Puzzle inspired me to try to go down that route.

worlds and the anchor that we keep fixed. Instead of evaluating a proposition at a world, we will evaluate it relative to a set of propositions: roughly, p is safe at w relative to F if for any nearby world v it is a member of F(v) or a member of F'(w). On this interpretation the set of propositions built out of F(v) and F'(w) represents an epistemically possible (but potentially metaphysically impossible) world, where the facts that have to do with the factor of interest F are like they are at v and all the non-F related facts are as they are at w. This is akin to Hawthorne's (2006) analysis of semantic plasticity: taking a part of a nearby world (the part responsible for language) and 'gluing' it to the actual world.

One might ask: why go through the trouble of constructing such impossible worlds by piecing together different parts of two worlds, instead of introducing them in a standard way as more or less arbitrary sets of sentences? The reason for my construction is that I want to preserve the accessibility relation on possible worlds: by using possible worlds as our building blocks we are able to make use of the already existing closeness relation on possible worlds.

For the purposes of building this model, I am going to assume that propositions are more fine-grained than on the simple model: here they are more structured, sentence-like entities. This is going to impact the way in which we carve up the worlds into our sets F(w) and F'(w). Consider the construction of F(w). F(w) contains the minimal proposition true in all the worlds of w's cell under the partition by F and all the propositions that this minimal proposition entails. However, a question arises as to what notion of 'entailment' is appropriate for our construction of F(w) and F'(w). On the one hand, the notion of metaphysical entailment (p metaphysically entails q iff $\Box(p \rightarrow q)$) is too strong. Consider again our two propositions p and q:

- p – the proposition that 'tallness' refers to the property of being over 180 cm in height
- q – a proposition expressing some precise description of the relevant usage facts such that $\Box(p \leftrightarrow q)$.

Even though p and q are necessarily equivalent, we would like to construct our sets so that relative to a factor like semantic plasticity, these propositions would be treated differently. On the other hand, the notion of logical entailment (p logically entails q iff q is a logical consequence of p) seems too weak. That is because we might have propositions that are problematically entangled with the subject matter F. Suppose that our subject matter F is how big some building B is. Intuitively, whether a building is big depends both on how tall it is and how wide it is; suppose that it is big if and only if it is both wide and tall. If our set F(w) contains the proposition that B is big, then it should also contain the propositions that B is wide and that B is tall. However, to achieve this we need something like an analytic entailment relation ($\Rightarrow$) that would tell us that x is big $\Leftrightarrow$ x is tall $\wedge$ x is wide. Therefore, I will use entailment in this sense: something stronger than the logical consequence relation, but weaker than metaphysical entailment.

Using this entailment relation we can construct our desired sets F(w) and F'(w). F(w) contains the minimal proposition of F(w) and all the propositions entailed by it. All the propositions that are in F(w) are to be removed from F'(w). Furthermore, we have to remove from F'(w) all the propositions that are at least partly about F, i.e. all propositions p* such that either p* or its negation is entailed by the minimal proposition in some F-cell.

Since we will not be evaluating propositions at possible worlds but rather at sets of propositions constructed from pieces of two worlds, we will need to close the constructed set under entailment. Let the angle brackets $\langle\rangle$ denote the closure under entailment of sets within the brackets, i.e. $\langle$F(v), F'(w)$\rangle$

denotes the closure under entailment of sets F(v) and F'(w).[77] The idea is that in case of propositions that are partly about our subject matter F, they will be neither in F(v) nor in F'(w); to recreate them, we need a closure over F(v) and F'(w). We will need an important assumption to deal with some problematic cases: if p is a proposition partly about F, then there are propositions q and r such that $p \Leftrightarrow q \wedge r$ and q is wholly about F and r wholly not about F. Let's call it the Analysability Constraint.

Let again our model be a quadruple M = ⟨W, R, $\mathcal{F}$, V⟩, where:

- W is a set of possible worlds,
- R is a reflexive accessibility relation,
- $\mathcal{F}$ is the set of factors: F ∈ $\mathcal{F}$ iff F is a partition of W (set of members of W),
- V is a function from worlds to sets of atomic propositions.

⟨F(v), F'(w)⟩ is the closure under entailment over the sets of propositions F(v) and F'(w), as outlined above. The interpretation function V gives us for each world w a set of atomic propositions true at w. The valuation is defined as follows (p stands for an atomic proposition and ϕ and ψ are schematic variables):

$\lvert p \rvert_w = 1$	iff	p ∈ V(w)
$\lvert \neg\phi \rvert_w = 1$	iff	$\lvert \phi \rvert_w = 0$
$\lvert \phi \wedge \psi \rvert_w = 1$	iff	$\lvert \phi \rvert_w = 1$ and $\lvert \psi \rvert_w = 1$

We can define our safety operator S as follows (ϕ is any sentence free of the safety operator):

$\lvert S(\phi, \mathrm{F}) \rvert_w = 1$	iff	for all v ∈ W such that $v\mathrm{R}w$:
		(i) $v \in c_F(w)$ and $\lvert \phi \rvert_w = 1$; OR

[77] ⟨F(v), F'(w)⟩ is defined as: $\left\{p: \exists q_i \ldots \exists q_n \left(\forall q_i \ldots \forall q_n \left(q_i \in F(v) \vee q_i \in F'(w) \right) \wedge \left(\bigwedge_{i=1}^{n} q_i \right) \Rightarrow p \right) \right\}$.

(ii) $\phi \in \langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$.[78]

The definition of S works just like on the simple model with the exception that ϕ is evaluated relative to the constructed set $\langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$ rather than a possible world.

Again, we can make similar observations about factor relative safety as depicted by simple model (although the proofs are slightly different in some cases).

OBSERVATION 1. S is factive: if $S(\phi, \mathrm{F})$ is true then ϕ is true.

OBSERVATION 2. S obeys Distribution: if $S(\phi, \mathrm{F})$ and $S(\phi \rightarrow \psi, \mathrm{F})$, then $S(\psi, \mathrm{F})$.[79]

OBSERVATION 3. If the subject matter F includes the subject matter G, then if $S(\phi, \mathrm{F})$ then $S(\phi, \mathrm{G})$.[80] The definition of subject matter inclusion is slightly different than before: F includes G iff for all worlds w and all propositions p, if $p \in \mathrm{G}(w)$ then $p \in \mathrm{F}(w)$.

---

[78] We can make S iterable in the same way as on the simple model.

[79] Proof is the same as on the simple model. Case 1: if F is general enough that all nearby worlds belong to the same cell, the proof is trivial: if ϕ and $\phi \rightarrow \psi$ are true in w, then ψ is too. Case 2: if for all nearby worlds v, $\phi \in \langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$ and $\phi \rightarrow \psi \in \langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$, then our closure condition guarantees that $\psi \in \langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$.

[80] Proof is slightly different than on the simple model. We have to prove that for some proposition p if for all nearby worlds v it is in $\langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$, then it is also in $\langle \mathrm{G}(v), \mathrm{G'}(w) \rangle$. If p is in $\mathrm{G'}(w)$ then there is no problem: any proposition that is in $\mathrm{G'}(w)$ is safe. Therefore, suppose that p is not in $\mathrm{G'}(w)$. There are two cases when p is not in $\mathrm{G'}(w)$. Case 1: p is wholly about G. If p is wholly about G, either p or its negation is entailed by every cell under partition by G. Therefore, for any nearby world v, either p or its negation is a member of $\mathrm{G}(v)$. Since, F includes G, any F-cell will entail either p or its negation. Therefore, for all the nearby worlds v, if $\mathrm{F}(v)$ contains p, then $\mathrm{G}(v)$ also contains it. Case 2: p is partly about G - there is a G-cell that entails either p or its negation. If this is the case then we can represent p as a conjunction r∧q (by the Analysability Constraint) such that r is wholly about G and q is neither partly nor wholly about G. For all nearby worlds v, if $\mathrm{F}(v)$ contains r, then $\mathrm{G}(v)$ also contains it (as argued in Case 1). Moreover, since q is neither partly nor wholly about G, q is in $\mathrm{G'}(w)$. Therefore, r∧q, and hence p, is in the closure over $\mathrm{G}(v)$ and $\mathrm{G'}(w)$. Consequently, if p is in $\langle \mathrm{F}(v), \mathrm{F'}(w) \rangle$, then it is also in $\langle \mathrm{G}(v), \mathrm{G'}(w) \rangle$.

OBSERVATION 4. $S(\phi, \text{F})$ and $S(\phi, \text{G})$ don't jointly entail $S(\phi, \text{F\&G})$.[81]
OBSERVATION 5. For any ϕ if for any subject matter F, $S(\phi, \text{F})$, then ϕ is safe *simpliciter*.[82]

We can use our complex model to deal with some more problematic cases of factor-relative safety. Suppose again that p is the proposition that 'tallness' refers to the property of being over 180 cm in height and q is a proposition expressing some precise description of the relevant usage facts such that $\Box(p \leftrightarrow q)$. Semantic plasticity prevents us from coming to know p, but not necessarily from coming to know q. Suppose that our factor F is the interpretation of the language. If p is true at w, then p is a member of $\text{F}(w)$. However, q will not necessarily be a member of $\text{F}(w)$, even though it is necessarily equivalent to p, because it need not be entailed (in our analytic sense of entailment) by the propositions contained in $\text{F}(w)$. Thus, the complex model allows us to handle some slightly more difficult problems than the simple model.

6 What to do with factor-relative safety?

The epistemicist employs a slogan: borderlineness is ignorance due to semantic plasticity. Since the epistemicist explains ignorance by failures of safety, I set out to explain the notion of safety (or lack thereof) relative to a factor. One natural expectation that one might have is that since we have our models for

[81] Proof is the same as on the simple model. Let F and G be two orthogonal factors and F&G, be the factor that includes both F and G. Suppose that $p \Leftrightarrow q \vee r$ such that q is wholly about F and r wholly about G. Furthermore, suppose that both r and q are true in w, but there is a nearby world w^* where both r and q are false. For any nearby world v, $\text{q}\vee\text{r}$ will be both in $\langle \text{F}(v), \text{F'}(w)\rangle$ as well as $\langle \text{G}(v), \text{G'}(w)\rangle$. Since r is wholly about G (which is orthogonal to F), r is in $\text{F'}(w)$ and consequently $\text{q}\vee\text{r}$ belongs to $\langle \text{F}(v), \text{F'}(w)\rangle$. By mirror reasoning, $\text{q}\vee\text{r}$ belongs to $\langle \text{G}(v), \text{G'}(w)\rangle$. However, under the partition by F&G, neither q nor r will remain in $\text{F\&G'}(w)$ and we know that there is a nearby world w^* where neither q nor r are true. Therefore, safety relative to F and safety relative to G, don't jointly entail safety relative to F&G.

[82] Same proof as on the simple model.

analysing factor relative safety in place, we ought to be able to derive an account of the definiteness operator Δ that would express the notion of safety relative to semantic plasticity. Unfortunately, this is not a straightforward task.

The notion of definiteness for the epistemicist concerns not only the proposition expressed by an utterance (or a belief), but also the sentence that is used to express the proposition. The standard explanation for why a belief that Michael is tall is not safe in a borderline case is that if one were to express that belief using the sentence 'Michael is tall', one could easily say something false due to semantic plasticity. However, the accounts of factor-relative safety developed above don't concern sentences used to express beliefs, they only concern the propositions believed. This issue illustrates two limitations of the account.

First, to give a full treatment of the epistemic notion of safety relative to semantic plasticity, we would need to include relativisation to methods in the model. It is a well-known feature of the safety-based accounts that they should include a relativisation to methods. Suppose that in the actual world I learn from a friend that they heard the announcement that train that I am about to take is delayed. It seems that I am in a position to know that the train will be delayed, even if in a nearby possible world I come to form the false belief the train will be delayed as a result of guessing that it will be delayed. The failure of my belief in a nearby world is does not constitute a relevant failure of safety, because the method I use in the nearby world where I make an error is different from the method I actually use. So to give a full account of factor relative safety relevant for epistemic purposes, we should amend the account to include a relativisation to safety.

Secondly, the epistemicist explanation of ignorance due to semantic plasticity involves a counterfactual false belief constituting a failure of safety

of the actual belief, despite the fact that the actual and the counterfactual belief have different contents. The reason that I am not in a position to know that Michael is tall, even though he is, is that in a close counterfactual case 'Michael is tall', as a result of semantic plasticity, expresses a slightly different (and false) proposition. Since the proposition believed in the counterfactual world would be slightly different than the proposition believed in the actual worlds, it seems that we need something like a counterpart relation on propositions. This relation would track which proposition would be believed by the agent had the meaning-facts been slightly different. Therefore, it seems that what we would need to handle such cases within the models of factor relative safety is adding a counterpart relation on propositions that would map agent's beliefs held in the actual world to the beliefs that would be held by the agent in counterfactual worlds (this again would have to be relativised to methods it seems).

These additions would significantly complicate the model. I don't think it would be impossible to develop a model with these features but doing this is beyond the scope of this chapter. If we are interested in developing an epistemicist account of definiteness, it might be better to tackle the problem more directly. This is the task for the next chapter: developing a definiteness operator for epistemicism.

7 Conclusion

My aim in this chapter was to shed some light on the epistemicist claim that what characterises definiteness is safety relative to semantic plasticity. I have provided two models to analyse factor-relative safety which help us to understand the notion better. However, the models of factor-relative safety are still quite basic: we would still need to introduce significant refinements for the

model to be applicable for the epistemicist needs. The task for the next chapter is to use the ideas developed here to define an epistemicist definiteness operator.

Chapter 5

Defining Definiteness

ABSTRACT

The recent literature (Caie 2011, Magidor 2018, Yli-Vakkuri 2016) points to a missing piece in the epistemicist theory of vagueness, namely a clear account of the semantics of the definiteness operator Δ. The fundamentals of the epistemicist theory are well understood. However, the technical work of defining the definiteness operator has proven difficult. In this chapter, I present a novel version of a multidimensional model and I provide an account of epistemicist definiteness that meets our desiderata.

WORD COUNT: 9,376

1 Introduction

The aim of this chapter is twofold. Firstly, to present a refined version of a Stalnakerian multidimensional model that would allow us to answer broader metasemantic questions. Secondly, to solve a technical problem in defining the definiteness operator for the epistemicist theory of vagueness.

Robert Stalnaker (1978) uses a multidimensional (2D) model to present his metasemantic theory and shows how the expressions that we use might have different content had the metasemantic context been different. He employs a

simple metasemantic context and a simple circumstance of evaluation; his analysis is concerned with how the changes in the metasemantic context change the content of expressions. On the other hand, David Kaplan (1977) uses a multidimensional model to account for the semantics of demonstratives and indexicals. I propose a multidimensional model, which squares the approaches of Kaplan and Stalnaker: a model that would allow us to handle indexicality (as Kaplan wants) and also answer metasemantic questions.

Furthermore, I make use of the model to solve an important problem for the epistemicist theory of vagueness (Williamson 1994b). The recent literature (Caie 2011, Magidor 2018, Yli-Vakkuri 2016) points to a missing piece in the epistemicist theory, namely a clear account of the semantics of the definiteness operator Δ. Attempts explored in the literature often employ a multidimensional model to provide the semantics of Δ; however, all these attempts face big obstacles. The multidimensional model I outline allows us to handle these problems and provide a clear semantics for the definiteness operator.

2 Multidimensional models

Since I will be using multidimensional models to present an account of the definiteness operator, it's worth to take the time to explain how my model will fit with the multidimensional accounts developed in the literature. The simple idea behind multidimensional models is that formulas in the language are not to be evaluated at a single world, but rather at more complex points of evaluation (e.g. a pair of worlds). This complication in the model allows us to account for some complexities in our language that we wouldn't otherwise be able to do. The majority of literature on multidimensional models uses two-dimensional frameworks, which use double-indexed points of evaluation

consisting of a context and circumstance of evaluation. Philosophers use multidimensional models for a variety of purposes. Two of the most important uses of multidimensional models come from the work of David Kaplan and Robert Stalnaker.

Kaplan (1977) uses two-dimensional models to account for the semantics of demonstratives and indexicals. On his account, formulas are assigned characters - functions from worlds to contents (intensions). The first index, the context of utterance, determines the content of the expression given by the character function. For instance, the character of 'I' is a function that for any context assigns the speaker in that context as the referent of 'I'. Contents are then evaluated relative to the second index (the circumstance of evaluation). Kaplan's account uses the two-dimensional model to explain the semantics of a certain class of expressions: a character of an expression is a certain kind of meaning.

On the other hand, Stalnaker (1978) uses the 2D model to show us how the expressions that we use might have meant something different than they actually do had the external world been slightly different. On this understanding of the 2D model, the function from the first index to intensions is not a kind of meaning like it is for Kaplan, intensions are the only kind of content there is[83], it merely shows what the content of our utterances might have been.[84] For instance, presumably the metasemantic rule for the term 'water' in our language is that it picks out a substance with chemical composition that is identical to the composition of the stuff that is found in rivers and lakes. In the actual world the chemical composition of the stuff

[83] Stalnaker refines his approach to handle indexicals in his later work. However, I am concerned here only with his basic insight about the interpretation of the 2D model.

[84] Another prominent interpretation of the 2D model is given by David Chalmers and Frank Jackson, i.e. the epistemic interpretation of the 2D model. I will not be concerned with this kind of interpretation in this chapter.

found in rivers and lakes is H_2O. However, in some different possible world *w**, the stuff in rivers in lakes is some compound XYZ - our metasemantic rule for 'water' would dictate that 'water' refers to XYZ in that world.

Despite the fact that Stalnaker and Kaplan give different interpretations to the 2D model we don't have to treat their approaches to be necessarily at odds: they simply employ the 2D model for different purposes. Stalnaker's interest is in his metasemantic project: he uses the 2D model to show how words in our language would have a different meaning if the external world had been different. On Stalnaker's model there is a simple metasemantic context, which determines the meanings of expressions in the language, and a simple circumstance of evaluation. Kaplan is interested in the analysis of indexicals. This means that he keeps the meanings (characters) of expressions in the language fixed. From a Stalnakerian perspective, Kaplan does not look at the effects of change in the metasemantic context (he keeps it fixed), but instead introduces a complex multi-dimensional circumstance of evaluation: expressions in the language are first evaluated relative to context, which tells us what intension should be assigned to expressions in our language, and then relative to a circumstance of evaluation (which gives us the semantic values).

Therefore, we could devise a model that attempts to square the approaches of Stalnaker and Kaplan: a model that allows us to answer metasemantic questions and is also able to handle indexicality. One way of doing this would be to introduce a three-dimensional model, on which we have a simple metasemantic context like Stalnaker, but have a double-indexed point of evaluation.[85] For each expression, the metasemantic context gives us its character, which allows us to evaluate the expressions relative to a context and

[85] This is the strategy adopted by Litland & Yli-Vakkuri (2016) and Yli-Vakkuri (2016).

circumstance of evaluation like on Kaplan's model. Therefore, on a three-dimensional model the function from the metasemantic context of utterance to characters is not a kind of meaning (analogously to the way that the 2D matrices don't represent a kind of meaning for Stalnaker). Such function merely tells us what character an expression has in different metasemantic contexts.

However, such a three-dimensional model will not be sufficient for my purposes. Just like substituting the unidimensional circumstance of evaluation for a double-indexed point is needed to handle some complex features of semantics like indexicality, we also need to substitute a simple metasemantic context for a double-indexed point to handle complexities in the metasemantics. Stalnaker's model is used to show how different metasemantic contexts would produce different meanings in our language. Stalnaker is concerned with the way that differences in the external world may affect the content of one's utterance. For instance, the metasemantic rules for assigning a semantic value for the term 'Hesperus' are such that 'Hesperus' is a rigid designator, referring to whichever celestial body is the brightest one in the evening. Actually, Venus is the brightest celestial body in the evening, so 'Hesperus' rigidly designates Venus. However, if there is a possible world w^* where the brightest celestial body in the evening is Mars then our metasemantic rules for assigning the content to 'Hesperus' would assign Mars as the reference of 'Hesperus'. So in sum, Stalnaker is concerned with exploring the way in which, by applying our metasemantic rules, the expressions in our language would get assigned different contents had the external world turned out differently.

My aim here is different. Whereas Stalnaker wants to keep the metasemantic rules (such as the rule that 'Hesperus' refers to the brightest celestial body in the evening) fixed, I am also interested in investigating cases where the metasemantic rules are different. For instance, we could be interested

in investigating how changing the metasemantic rules while keeping the external world fixed would affect the contents of our language. On this approach, the question regarding 'Hesperus' would be: had our metasemantic rules been different, how what would be the content assigned to 'Hesperus'? For instance, if in a world w^* the metasemantic rules regarding 'Hesperus' are such that it is a rigid designator that refers to the Moon, then by the metasemantic rules for 'Hesperus' refers to the Moon in the actual world.

For this reason, we need a double-indexed metasemantic context. The first index gives us the metasemantic rules that govern the expressions in the language (such as the rule that 'Hesperus' refers to the brightest celestial body in the evening): it gives us the function from worlds to characters. The second index gives us what we may dub the *metasemantic circumstance of evaluation*: it provides an input to the function determined by the first index. For instance, if we apply the actual metasemantic rule for 'Hesperus' to a world w^* where Mars is the brightest celestial body, we get a constant character picking out Mars.

My aim in what follows will be to outline the epistemicist theory of vagueness and present a multidimensional model that would allow us to define a definiteness operator for epistemicism. In Section 3, I outline the epistemicist theory of vagueness and I present my understanding of epistemicist definiteness. In Section 4, I present the desiderata for defining definiteness. In Section 5, I present a multidimensional model where I define definiteness and show that this account can deal with some problematic cases.

3 Safety and definiteness

The epistemicist project uses safety-based epistemology to explain the ignorance that arises in borderline cases. On the safety conception, knowledge

is incompatible with a possibility of an easy error. The epistemicist identifies a feature of vague expressions in our language, namely *semantic plasticity*, and argues that semantic plasticity gives rise to a possibility of easy errors in borderline cases. An expression is semantically plastic if and only if small shifts in the use of language across close possible worlds unbeknownst to us slightly shifts the intension of the expression. It's easy to see how semantic plasticity could give rise to ignorance in borderline cases. If Michael is a borderline case of 'tallness', then even if Michael is *tall*, he is not *tall** (where *tallness** is a slightly different property from *tallness* such that we could easily refer to it had used the language slightly differently). This means that if I express my belief that Michael is tall using the sentence 'Michael is tall', there is a possibility for me to make an easy error: if we used the term 'tall' so that it refers to *tallness** (and it could easily be the case that we use the language that way), 'Michael is tall' would say something false.

What is definiteness on the epistemicist account? If ignorance in borderline is associated with the easy possibility of error due to semantic plasticity, then definite cases are ones where semantic plasticity does not pose such a threat to safety. This does not mean that definite cases are ones that are *overall* safe, it only means that they are safe relative to semantic plasticity (as an ignorance inducing factor). This is to be expected; after all, 'definitely p' does not entail that we are in a position to know p, just that vagueness alone cannot prevent us from being in a position to know p. Therefore, on the epistemic treatment of vagueness, it's definitely ϕ when the variation in semantic facts (facts having to do with what intensions are assigned to expressions in our language) is insufficient to produce an error across the space of nearby worlds.

There is a question about how important it is for the epistemicist to give an interpretation of the definiteness operator. On the Non-Ambitious reading of epistemicism discussed in Chapter 4, this could be seen as a superfluous

addition. One of the important advantages of epistemicism is its parsimony. The theory employs (a) the independently established safety-based account of knowledge and (b) the postulate that expressions in our language are semantically plastic, which is more controversial but also independently motivated (Dorr & Hawthorne 2014). Using these two pieces of theoretical machinery, the epistemicist wants to explain, fundamentally, why there is ignorance in borderline cases. If the epistemicist is able to explain ignorance in borderline cases, why is there a need for additional category of definiteness? If a proposition is definitely true, it just means that it's safe relative to semantic plasticity. However, semantic plasticity is one of many ignorance inducing factors. I could be ignorant as to whether Michael is tall because 'Michael is tall' is semantically plastic, but I could also be ignorant because I cannot see Michael very well from a distance. It's not clear that the epistemicist should treat *safety relative to semantic plasticity* as a particularly important category as compared to safety relative to imperfect eyesight.

However, going forward I will take the Ambitious Epistemicist route and assume that defining the definiteness operator is a worthwhile task for the epistemicist. There are certainly dialectical advantages to having such an interpretation, not the least of which is being able to paraphrase the discourse employed by other theories of vagueness (such as the supervaluationist or metaphysical accounts) for whom definiteness does constitute a special category. If we are able to provide a clear interpretation of the definiteness operator, there is no harm in doing so.

4 Desiderata for an account of definiteness

There are several desiderata that the epistemicist definiteness operator should meet. The accounts presented in the literature so far struggle to meet one or

more of these criteria. My aim is to outline these desiderata and show that my account meets them. Then, in Section 6, I will show that alternative formulations of the epistemicist notion of definiteness fail to meet the desiderata.

4.1 Isolation of semantic plasticity

The epistemicist associates borderlineness with ignorance produced by semantic plasticity. Recall the Definiteness Puzzle discussed in the previous chapter:

> (Definiteness Puzzle). Suppose that Michael is 190 cm in height. The actual cut-off point for 'tall' is 185 cm in height and the margin for error is 3 cm. Since Michael's height is not within the margin for error, Michael counts as being definitely tall. Consider a world w^* where Michael is 3 cm shorter and where the cut-off point for 'tall' is 188 cm. Such a world should count as close: Michael is only slightly shorter than in the actual world and the cut-off point is within the margin of error. However, Michael is not in the extension of 'tall' at w^* as it is used at w^*. So Michael cannot be definitely tall in the actual world, because there is a close possible world where 'Michael is tall' is false at a close possible world; thus, a belief that Michael is tall is not safe and does not constitute knowledge.

It is clear that the falsity at w^* of 'Michael is tall' as used at w^* is due to both the variation in non-semantic facts (that have to do with Michael's height) and the semantic facts (that have to do with the boundary for 'tall') between the actual world and w^*. This is not the kind of variation that we are interested in. We are interested to know whether in a given case it is semantic plasticity alone could produce an easy error. The epistemicist account of definiteness should be formulated as to isolate the effect of semantic plasticity.

As discussed in Chapter 4, since semantic plasticity is variation in semantic facts what we would ideally like to do is to keep the non-semantic facts fixed and only allow for variation in the semantic facts to isolate semantic plasticity as the ignorance inducing factor. If it were possible, we would like to evaluate sentences at nearby worlds which differ from the actual world at most with regards to the semantic facts (but not with regards to the non-semantic facts). Unfortunately, this is not possible: the semantic facts supervene on the non-semantic facts.[86] Keeping the non-semantic facts fixed would also have the unintended consequence of keeping the semantic facts fixed. Therefore, we need another approach.

The Definiteness Puzzle allows us to make the following observation: a definition of Δ should incorporate the insight that '$\Delta\phi$' ought to be false only if variation in the intension of ϕ alone could easily make ϕ false.

4.2 Logic of Δ

The definiteness operator should satisfy the typical constraints.

(Factivity) $\Delta\phi \rightarrow \phi$

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

(Intransitivity) $\neg\forall\phi(\Delta\phi \rightarrow \Delta\Delta\phi)$

(Taut.) For all logical tautologies ϕ, $\Delta\phi$ is true.

(Necessitation) For all ϕ such that $\vdash \phi$, $\Delta\phi$ is true.

[86] Kearns & Magidor (2008) argue that semantic facts are sovereign: they don't supervent on the non-semantic facts. Some additional avenues of defining definiteness could be available to an epistemicist who embraces semantic sovereignty. However, I will proceed via a more traditional route that accepts the supervenience of the semantic on the non-semantic.

4.3 Interaction with modal operators

When giving semantics for Δ we should be careful to define an operator that interacts well with the metaphysical modality operators. There are cases where our definiteness talk interacts with modal talk. For instance, even though I am actually definitely not bald, it's possible for me to be definitely bald. There are certain conditions we would like the definiteness operator to satisfy. For instance, on the correct definition of Δ the following should come out as true:

(i) $\Box(\Delta\phi \rightarrow \phi)$

On the other hand, the following principles should not be universally true:

(ii) $\Box(\Delta\phi \vee \Delta\neg\phi)$

(iii) $\Delta\phi \rightarrow \Box\Delta\phi$

(iv) $\Box\phi \rightarrow \Delta\phi$

(v) $\Delta\phi \rightarrow \Box\phi$

4.4 Knowability Constraint

The defined operator should have an appropriate relation to knowledge – if we are in a position to know ϕ then $\Delta\phi$. In particular, the following disquotational schemas should come out as definite:

(T-Schema) 'ϕ' is true $\leftrightarrow \phi$

(R-Schema) 'τ' refers to τ

The task for the epistemicist is to find an account of definiteness that would satisfy all these desiderata.

5 Epistemicist Definiteness

5.1 How to build a model of epistemicist definiteness?

My approach to defining the epistemicist definiteness operator mainly draws inspiration from three sources: Stalnaker's metasemantic interpretation of multidimensional semantics, John Hawthorne's (2006) theoretical approach to defining epistemicist definiteness and Juhani Yli-Vakkuri's (2016)[87] technical work on defining definiteness. The plan is to use Hawthorne's insights regarding the understanding of definiteness in an epistemicist setting and use a multidimensional model to define Δ.

John Hawthorne (2006) provides a good illustration of what we should be after in defining the epistemicist notion of definiteness. The epistemicist claims that in order for ϕ to be definitely true, we need to evaluate ϕ relative to how it is used in all nearby worlds. Hawthorne (2006) illustrates this by imagining that a speaker of a non-actual close variant of the language (a Twinglish speaker) is with us in the actual world and we are evaluating his utterances made in the actual world. This does not mean that there are actually speakers of non-actual variants of English in the actual world. However, Hawthorne's picture gives us a good intuitive understanding of what the epistemicist is after. By pretending that there is a community of Twinglish speakers in the actual world, we are able to isolate semantic plasticity as an ignorance inducing factor. We keep the worldly (non-semantic) facts fixed and let the language vary which allows us to determine whether a shift in the linguistic facts alone is able to produce an easy error. Consider Michael's case from the Definiteness Puzzle. Michael is 190 cm in height and in English the boundary for 'tall' (185 cm) is

[87] See also Litland & Yli-Vakkuri (2016).

below that height, so he counts as tall. In Twinglish the boundary for 'tall' is 188 cm. Had the Twinglish speaker been with us in the actual world, his utterance 'Michael is tall' would be true as Michael's actual height is over the boundary.

Using the terminology I employed in Section 2, we may say that the epistemicist is interested in discovering how changing the metasemantic rules of our language would impact its semantics (while keeping the metasemantic circumstance of evaluation fixed). One might naturally wonder why the epistemicist needs a double-indexed rather than a simple metasemantic context. This need is demonstrated by the following case. Suppose that that there is a lottery with 1000 tickets and only one winner. Let the name 'Chancy' denote the winner of the lottery. Even if I don't know which ticket won the lottery, it seems that I am still in a position to know that Chancy is the winner of the lottery, because I have introduced the name 'Chancy' so that it refers to the winner.[88] Even if another ticket would have won, I would be in a position to know that Chancy is the winner, because then 'Chancy' would refer to that (other) winner. If I know that Chancy is the winner, then it's definitely the case that Chancy is the winner. However, if we go with a simple metasemantic context that determines the intension of 'Chancy' we will get the wrong result. That is because in some close world w^*, some ticket other than Chancy, let's call it 'Schmancy', is the winner. Therefore, 'Chancy' as uttered in the

[88] This is consistent with what we would want to say about Chancy's case in terms of the definiteness status of 'Chancy is the winner': we would want to say that it is definitely the case that Chancy is the winner. Furthermore, we would want to say of Chancy (i.e. *de re*) that it is definitely the winner as there is no vagueness as to which of the tickets 'Chancy' refers to. However, when talking about knowledge simpliciter, we would say that we know that Chancy is the winner, but we would not say *de re* of Chancy that we know that he is the winner (for all we know some other ticket won). The difference in the case of *de re* vagueness and *de re* knowledge is important: we want to know how semantic plasticity in particular produces ignorance and not which propositions are known in general.

metasemantic context w^* will refer to Schmancy. But Schmancy is not the winner in the actual world (Chancy is). Therefore, we need a double-indexed metasemantic context: we change the first index, which gives us the metasemantic rules for 'Chancy', but we keep the metasemantic circumstance fixed.

The epistemicists are interested in the effect that change in metasemantic rules that would have on the reference of 'Chancy' while keeping the external facts (such as which ticket is the winner) constant. What we need instead is an interpretation that would mimic Hawthorne's setup. If the Twinglish speaker were with us in the actual world, their utterance 'Chancy is the winner' would be true, because 'Chancy' in Twinglish would refer to Chancy. The reason for this is that we have a clear rule of determining the intension of 'Chancy': it picks out the actual winner of the lottery. Since we have a clear rule of determining the intension of 'Chancy', we can expect that this rule applies to every close variant of the language. Furthermore, if the Twinglish speaker were with us in the actual world, then 'Chancy' in Twinglish would refer to Chancy, because there is only one winner of the lottery in the actual world, and it is Chancy. Consequently, instead of letting a simple metasemantic context determine the intension (or character) of 'Chancy', we need something slightly more complicated.

The setup proposed by Hawthorne requires a complex metasemantic context. Firstly, we need a metasemantic function that would tell us how to assign characters to expressions given the facts in the external world. Secondly, we need a world relative to which the characters are to be assigned to expressions. We need these two elements of metasemantic context to explain why it's definitely the case that Chancy is the winner of the lottery. In case of 'Chancy', the first element of metasemantic context gives us the rule that tells us that the winner of the lottery is the referent of 'Chancy' and the second

element gives us the world from which the winner is drawn. It's definitely the case that Chancy is the winner, because (a) all nearby variants of the language give us the same metasemantic rule that assigns the winner of the lottery as the intended referent of 'Chancy' and (b) we keep the second element of the metasemantic context fixed on the actual world. This is essentially what Hawthorne does when he imagines that the Twinglish speaker is with us in the actual world: the first element of metasemantic context changes (the element that gives us the function from worlds to characters), but the second element of metasemantic context is kept fixed (so the external facts which determine meanings/characters in our language are kept fixed).

Switching from a Stalnakerian 2D model to model with a complex circumstance of evaluation (a pair of worlds) allows us to handle indexicality. Furthermore, what is needed for handling Hawthorne's approach to definiteness within the multi-dimensional model is a switch from a simple metasemantic context (a single world) to a complex one (a pair of worlds). The first element in the pair gives use the 'descriptive' metasemantic rule that tells us how the facts in the external world determine the characters of expressions in the variant of the language, i.e. it gives us a function from worlds to characters. The second element is the external world that is an argument in the function. For instance, in the example of 'Chancy', the first element would give us the rule that the constant character of 'Chancy' picks out the winner of the lottery; the second element gives us the winner of the lottery (and determines which ticket in particular 'Chancy' will refer to).

Therefore, what we need is a multi-dimensional model with quadruple-indexed points of evaluation $\langle w_1, w_2, w_3, w_4 \rangle$ where:

- the first index, the *metasemantic context*, gives us the function from worlds to characters for the expressions in the language

- the second index, the *metasemantic circumstance*, gives us the character for the expressions in the language
- the third index, the *semantic context* (or context simpliciter), gives us the intensions
- the fourth index, the *circumstance of evaluation*, gives us the semantic values

In the next section I briefly outline the 4D model.

5.2 The 4D model

In the 4D model, formulas are evaluated relative to quadruple-indexed points of evaluation $\langle w, v, u, z\rangle$. Intuitively we may think about this as a refinement of a two-dimensional model where the simple metasemantic context and circumstance of evaluation is replaced by a pair of worlds (to handle metasemantic and semantic complications outlined above). The first index is the metasemantic context, which determines the variant of the language relative to which we are to evaluate formulas. The metasemantic context is something that the definiteness operator operates on, which allows us to see how small shifts in the interpretation (metasemantic rules) of an expression shift the semantics in the language.

A 4D-model is a triple $M = \langle W, R, [\![.]\!]\rangle$, where W is a set of worlds, R a reflexive accessibility relation[89] and $[\![.]\!]$ is a function from atomic sentences to sets of four-dimensional points (subsets of $W \times W \times W \times W$). The bivalent

[89] The accessibility relation R should be interpreted as the epistemic 'closeness' relation: w is R-related to v if w is close to (epistemically indiscriminable from) v. For simplicity, I assume that the accessibility relation relevant to the operation of standard metaphysical modality operators is universal, so I will use quantification over worlds (universal for necessity and existential for possibility) to give the interpretation to modal operators.

valuation of sentences at 4D-points of evaluation $\langle w,\ v,\ u,\ z\rangle$ is defined as follows (P stands for an atomic sentence and ϕ and ψ are schematic variables):

$\lvert\mathrm{P}\rvert_{w,\ v,\ u,\ z} = 1$	iff	$\langle w,\ v,\ u,\ z\rangle \in [\![\mathrm{P}]\!]$
$\lvert\neg\phi\rvert_{w,\ v,\ u,\ z} = 1$	iff	$\lvert\phi\rvert_{w,\ v,\ u,\ z} = 0$
$\lvert\phi \wedge \psi\rvert_{w,\ v,\ u,\ z} = 1$	iff	$\lvert\phi\rvert_{w,\ v,\ u,\ z} = 1$ and $\lvert\psi\rvert_{w,\ v,\ u,\ z} = 1$
$\lvert\Box\phi\rvert_{w,\ v,\ u,\ z} = 1$	iff	for all $z^* \in \mathrm{W}$, $\lvert\phi\rvert_{w,\ v,\ u,\ z^*} = 1$
$\lvert A\phi\rvert_{w,\ v,\ u,\ z} = 1$	iff	$\lvert\phi\rvert_{w,\ v,\ u,\ u} = 1$
$\lvert\Delta\phi\rvert_{w,\ v,\ u,\ z} = 1$	iff	for all $w^*\mathrm{R}w$, $\lvert\phi\rvert_{w^*,\ v,\ u,\ z} = 1$

We can say that ϕ is true *simpliciter* in w if $\lvert\phi\rvert_{w,\ w,\ w,\ w} = 1$. Logical consequence is defined as truth preservation in all proper points of evaluation (points of the form $\langle w,\ w,\ w,\ w\rangle$) in all models.

5.3 Satisfaction of desiderata

We are mostly interested in the definition of Δ. We have given the following truth conditions for Δ:

> (Epistemicist Definiteness) $\lvert\Delta\phi\rvert_{w,\ v,\ u,\ u^*} = 1$ iff for all $w^*\mathrm{R}w$, $\lvert\phi\rvert_{w^*,\ v,\ u,\ u^*} = 1$

The rationale for the semantics is as follows. We keep the metasemantic circumstance, the context, and the circumstance of evaluation fixed, and we vary only the metasemantic context, because we are interested in the change that would result from using different metasemantic rules to assign characters to our formulas. To borrow a term from Kaplan (1977), the definiteness operator is a metasemantic monster[90] on my view. We allow the metasemantic

[90] According to Kaplan (1977) monsters are operators that operate on contexts of utterances. Kaplan claims that English does not actually have monstrous operators, but that they nevertheless could exist in some other (perhaps merely possible) languages.

context (the first index) to shift across the space of close possible worlds. This allows us to see how the shift in usage across close linguistic communities influences the semantic value of expressions in our language.

The account has a few initial advantages going for it. Firstly, it does not fall prey to the Definiteness Puzzle: the account keeps the 'worldly' facts fixed as Epistemicist Definiteness (ED) does not shift the circumstance of evaluation. This allows us to isolate semantic plasticity as an ignorance-inducing factor: if the truth value of ϕ changes as we change the metasemantic context, then we can say that shifts in language can produce an easy error (and consequently ignorance). This means that ED meets the first desideratum for an epistemicist account of definiteness: isolation of semantic plasticity as an ignorance inducing factor.

Secondly, ED allows for a very natural interaction between the definiteness and metaphysical modality operators, and it validates the formulas using the necessity operator that we would intuitively want validated. For instance, our semantics validates the intuitive formula $\Box(\Delta\phi \rightarrow \phi)$. Furthermore, it correctly does not validate $\Box(\Delta\phi \vee \Delta\neg\phi)$ nor $\Delta\phi \rightarrow \Box\Delta\phi$ nor $\Box\phi \rightarrow \Delta\phi$ nor $\Delta\phi \rightarrow \Box\phi$.

Thirdly, the account allows us to handle important problematic cases that posed problems for some accounts presented in the literature. For instance, consider the sentence 'definitely, Chancy is the winner of the lottery'. This is validated by the account. It's definitely the case that Chancy is the winner, because the metasemantic rule that says that the ticket that wins the lottery is the same in all nearby worlds. Therefore, for all nearby worlds w, 'Chancy is the winner of the lottery' is true as evaluated at $\langle w, @, @, @\rangle$.

Furthermore, let's consider the case of disquotational sentences like:

(1) 'Tallness' refers to tallness.

Generally we want sentences like (1) to come out as definitely true. Suppose that (1) is uttered at $\langle @, @, @, @\rangle$ and we want to check whether it is definitely

true. We know that "Δ 'tallness' refers to tallness" is true at ⟨@, @, @, @⟩ only if for all close worlds w, "'tallness' refers to tallness" is true at ⟨w, @, @, @⟩. Whether (1) is definitely true will depend on our interpretation of metalinguistic vocabulary such as 'refers'. For instance, (1) will not be definitely true, if we interpret the utterance of (1) at a point ⟨w, @, @, @⟩ as if it were the case that a speaker from the metasemantic context w is talking about the actual variant of the language (because @ is the circumstance of evaluation, so (1) seems to be about how things are at @). As uttered in w, 'tallness' refers to *tallness** (some property distinct from *tallness*), but in the actual variant of the language 'tallness' refers to *tallness*. So if the speaker from w were talking about the actual variant of the language, they would be saying something false (as 'tallness' doesn't refer to *tallness** in the actual variant, it refers to *tallness*).

However, I don't think that interpreting (1) in this way makes much sense. (1) seems to be definitely true because no matter which close variant of the language we use to utter (1) it should come out as true; no matter what exact property is referred to by 'tallness', the fact that we use the word 'tallness' to talk about the property that 'tallness' refers to, guarantees that we are telling the truth. However, there is no intuition that says that no matter what variant of the language we use to evaluate the actual variant of the language '"tallness" refers to tallness' should come out as true. Suppose that we apply Hawthorne's setup to analysing definiteness to the case. If there were a speaker of the language coming from w^* (Twinglish) in the actual world, their Twinglish utterance '"tallness" refers to tallness' made in the actual would be about Twinglish not English. By uttering (1) we are not using any expressions identifying English specifically as the language about which the sentence is uttered. Rather, we are using the fact that naturally (1) is interpreted as being about the language variant in which the sentence is made. In other words,

when we ask whether "'tallness' refers to tallness" is definitely true, we are not looking to find out what speakers of other variants of the language would say about us. Rather, we want to know whether if we had been using a different language variant to talk about ourselves in that variant would what we are saying be true.

Therefore, what we want is for (1) as evaluated relative to $\langle w^*, @, @, @\rangle$ to be true. We can make this happen by assigning the appropriate truth conditions to metalinguistic vocabulary such as 'true' and 'refers'. This is a solution suggested by Juhani Yli-Vakkuri (2016, p. 822-823) to an analogous problem for his model of definiteness. We can stipulate that sentences of the form '"τ" refers to τ' and '"ϕ" is true if and only if ϕ' are true whenever the metasemantic circumstance, context, and circumstance of evaluation are identical. Suppose that we call the function from worlds to characters a *metasemantic character*. Furthermore, call a metasemantic character f *diagonally true* if and only if for all $w \in$ W, $\mathrm{f}(w)(w)(w) = 1$. The proposal is that in all the nearby worlds instances of disquotational schemas like '"τ" refers to τ' and '"ϕ" is true if and only if ϕ' have diagonally true metasemantic characters.[91] This would make them definitely true.[92]

---

[91] One issue that remains is what exact semantics should be given to metalinguistic vocabulary such as 'true' and 'refers'. On the one hand we want to validate sentences like '"τ" refers to τ', so we want diagonally true metasemantic characters. Consequently, if w^* is a nearby world we want '"τ" refers to τ' to be evaluated as true at $\langle w^*, @, @, @\rangle$. However, we also want our metalinguistic vocabulary to be sufficiently expressive to be able to talk about other language variants. This is not a problem. We can introduce additional vocabulary into the language that would allow us to talk about other language variants by referring to them specifically. For instance, we can talk about the language in some world w and say things like '"τ" refers to x at w'. The details of the treatment of metalinguistic vocabulary like 'true' and 'refers' requires a separate project.

[92] This is not an ad hoc solution. The metasemantic characters of '"τ" refers to τ' and '"ϕ" is true if and only if ϕ' are diagonally true because when we evaluate them at a point $\langle w^*, w, w, w\rangle$ (where w^* is a nearby world), we interpret the situation alongside

In the next section, I will outline alternatives to my account that have been presented in the literature. I will argue that each of these accounts suffers from issues that my account does not.

6 Alternatives and their issues

6.1 Simple Diagonal Account

The epistemicist wants ultimately to explain the ignorance due to vagueness by the failure of the safety principle. The basic idea is that even if some borderline sentence ϕ is true in the variant of the language we are actually speaking, it could easily be the case that we would be speaking a very similar variant of the language on which ϕ is false. We cannot discriminate between very similar variants of the language, so we could easily make an error, e.g. by asserting ϕ. This possibility of error is something that stands in the way of our knowledge. There is a natural proposal of defining definiteness in the light of the above remarks:

> (Simple Diagonal Account) ϕ is definitely true (at @) if and only if for all close worlds w, ϕ as used at w is true at w.

Unfortunately, there is a simple problem with this account: it does not isolate the effect of semantic plasticity as an ignorance inducing factor. Suppose that ϕ is actually true but at some nearby world w^*, ϕ (as used at w^*) is false at w^*. However, the account does not tell us why ϕ is false at w^*. For all we know it could be false because of factors other than semantic plasticity.

Furthermore, there are problems with the interaction of definiteness and metaphysical modality on the Simple Diagonal Account. Suppose that we are

---

Hawthorne's picture: we imagine a speaker of the variant of the language from w^* who is in w uttering the sentence.

evaluating a sentence 'I could be definitely bald'. This means that there is a world w^* such that 'definitely I am bald' is true as evaluated at w^*. However, in fact I am definitely not bald. If we were to evaluate the sentence 'I am bald' so that for all worlds w (such that w is close to @), 'I am bald' as uttered at w is to be evaluated at w, then 'definitely I am bald' would come out as false (since I am actually definitely not bald). However, if we were instead to evaluate the sentence 'I am bald' relative to all the worlds that are close to w^*, we also get the wrong result as 'bald' as uttered at w^* may mean something completely different from what it actually means (w^* could be any world for all we know, does not have to be a close one).

Therefore, the Simple Diagonal Account does not mean our desired desiderata: it does not isolate the effect of semantic plasticity as an ignorance inducing factor and does not interact well with metaphysical modality operators.

6.2 Complex Diagonal Account

Suppose that we define definiteness like this:

(Diagonal) $|\Delta\phi|_{w,\, v,\, u,\, z} = 1$ iff for all $w^* \mathrm{R} w$, $|\phi|_{w^*,\, w^*,\, w^*,\, w^*} = 1$

Unfortunately, there are multiple problems with the definition of Δ. One such problem is that on the above definition, metasemantic circumstance (second index), the context (third index) and world of evaluation (fourth) index are irrelevant to the evaluation of $\Delta\phi$ at some 4D point $\langle w,\, v,\, u,\, z\rangle$, i.e. no matter what worlds v, u and z are, the truth value of $\Delta\phi$ will depend only on the truth value of ϕ in worlds close to w (the metasemantic context). This is of course undesirable: it matters a great deal to the truth value of 'it's definitely the case that Napoleon lost the battle of Waterloo' whether the world of evaluation is one where Napoleon lost the battle or one where he won it.

We could of course revise the account. For instance, on a natural revision Δ could be defined as follows:

(Diagonal*) $|\Delta\phi|_{w,\, v,\, u,\, z} = 1$ iff for all z^*Rz, $|\phi|_{z^*,\, z^*,\, z^*,\, z^*} = 1$

On this definition, the initial world of evaluation does matter to the truth value of $\Delta\phi$, but other problems emerge. For instance, the initial context of utterance (second index) is still irrelevant to the evaluation of the sentence. This also clearly problematic, as it makes intuitively knowable (and therefore definitely true) propositions borderline. For instance, $A\phi \leftrightarrow \phi$ seems to be definite but turns out to be false on the revised account. Similarly, the above account falsifies the intuitively true sentence $\Box(\Delta\phi \rightarrow \phi)$. For suppose that ϕ is the sentence 'Hesperus ≠ Phosphorus'. It is necessarily the case that Hesperus is identical to Phosphorus (both are identical to the planet Venus) so for all z^*, $|\text{Hesperus} \neq \text{Phosphorus}|_{@,@,@,z^*}=0$. However, there are worlds (perhaps far away ones) where 'Hesperus' is used to refer to Mars and Phosphorus is used to refer to Venus. Suppose that w^* is such a world and additionally that the use of 'Phosphorus' and 'Hesperus' at all the worlds close to w^* is the same as in w^*. This means that $|\Delta\phi|_{@,@,@,w^*}=1$. However, it is still the case that $|\phi|_{@,@,@,w^*}=0$ because ϕ relative to the actual variant of the language still expresses the proposition that Venus is distinct from Venus. So the conditional $\Delta\phi \rightarrow \phi$ comes out as false.

The general technical problem with the diagonal strategy is that the definiteness operator is given the power to shift the context of utterance and the world of evaluation. The problem arises when we attempt to use both the definiteness operator and the modal operators in the same sentence as the competences of the different operators conflict. Having multiple operators that shift the same index is not always problematic of course, e.g. in case of multiple temporal operators, when it is clear that these different operators really should operate alongside the same dimension (e.g. time). However, in the case of

definiteness, there is no reason why the definiteness operator should have the power to shift the world of evaluation since definiteness is not a metaphysical modality operator.

6.3 Ground Fixing Account

Ofra Magidor (2018) has proposed an interesting account of definiteness that makes use of the notion of grounding. Magidor's proposal is that 'definite' should be given the following truth conditions:

> (Ground Fixing Account) Let p be the proposition that is expressed by ϕ at @ and let Q be the ultimate grounds for p in @. ϕ is definitely true at @ if and only if for all nearby worlds w where Q holds, ϕ as used at w is true at w.

This proposal is intuitively plausible. What went wrong with the Simple Diagonal Account, as demonstrated by the Definiteness Puzzle, is that we allowed the non-semantic facts such as ones related to Michael's height to vary across the set of close possible worlds. However, on Magidor's proposal the relevant non-semantic facts should be kept fixed. Facts about Michael's height are what grounds the truth of the sentence 'Michael is tall'; thus, on Magidor's proposal we only consider close worlds where the grounding facts that are relevant to the sentence 'Michael is tall' are kept fixed. This solution allows to isolate semantic plasticity as an ignorance inducing factor and avoid the issues brought to light by the Definiteness Puzzle.

Magidor's proposal has no problem in dealing with cases such as:

(1) 'Tallness' refers to tallness.

which correctly comes out as definitely true. However, the problem for the proposal is that it does not treat as relevant certain worlds that intuitively ought to count as relevant. This is visible in examples such as:

(2) 'Tallness' refers to the property of being over 185 cm height.

which comes out as definitely true on Magidor's proposal. By supposition in the actual world 'tallness' refers to the property of being over 185 cm in height. On Magidor's proposal, when evaluating the definite status of (2), we should only look at close possible worlds where the grounding facts relevant to the truth of (2) are the same as in the actual world. However, if all the facts that ground the truth of (2) are kept fixed across the relevant set of worlds, then (2) is true in all these worlds. If we only look at worlds where the grounding facts for the truth of '"tallness" refers to the property of being over 185 cm in height' are the same as in the actual world, then 'tallness' refers at these worlds to the same property as in the actual world, namely the property of being over 185 cm in height. So again (2) is definitely true on Magidor's proposal.

Magidor (2018) is happy to accept this consequence.[93] However, there are good reasons to think that Magidor's proposal doesn't pick out the notion of definiteness the epistemicist is after. What the epistemicist is after is the notion of safety relative to semantic plasticity. Is semantic plasticity an obstacle in coming to know that (2) is true? It is, but it's not the semantic plasticity of (2) that is responsible for our ignorance, it's the semantic plasticity of 'tallness'. Magidor's approach assumes that if semantic plasticity is an obstacle to knowing that ϕ is true, then it is the semantic plasticity of ϕ in particular that is responsible. However, it may be the case that the semantic plasticity of some other piece of vocabulary is responsible for our ignorance of the truth value of ϕ (e.g. an expression merely mentioned in ϕ). We need a global notion of safety

[93] An interesting consequence of Magidor's proposal is that the definiteness operator would fail to satisfy Distribution: $\Delta(\phi \to \psi) \to (\Delta\phi \to \Delta\psi)$. For instance, if we substitute (2) for ϕ and "'tallness' is the property of being over 185 cm in height" for ψ, then $\Delta(\phi \to \psi)$ and $\Delta\phi$ are true but $\Delta\psi$ is false.

relative to semantic plasticity that would look at the variation in the interpretation of a language as a whole.

To put the point another way, the reason for this is that the Ground Fixing Account goes against the spirit of the epistemicist view, because it does not treat as relevant worlds that intuitively should count as such. Expressions in our language are semantically plastic according to the epistemicist, which means that we are not sensitive to slight shifts in reference between close possible variants of the language. That is to say that we cannot discriminate between variants of the language that differ only slightly in the assignment of semantic values to the expressions. Naturally this means that worlds which differ slightly, but differ nevertheless, from the actual world with respect to what property the word 'tallness' refers to should count as relevant for our analysis, because these worlds are semantically indiscriminable from the actual world. However, on Magidor's proposal such worlds are not taken to be relevant, because the grounding facts for the truth of (2) are different there. Magidor's proposal is a good attempt at solving the problem at hand, but unfortunately it fails.

6.4 3D Model

Litland and Yli-Vakkuri (2016) develop a three-dimensional model to provide an account of definiteness. On the 3D semantics, formulas are evaluated relative to triple-indexed points of evaluation $\langle w, v, u \rangle$. The first index is the metasemantic context, which determines the variant of the language relative to which we are to evaluate formulas; the second index is the context of utterance, and the third index is the circumstance of evaluation. In short, instead of having a complex metasemantic point like my 4D model employs, Litland and Yli-Vakkuri (2016) have a single metasemantic context. The

second and third index are used in the usual way: they allow us to handle indexical expressions in a Kaplanian way.

A 3D-model is a triple M = ⟨W, R, [[.]]⟩, where as usual W is a set of worlds, R a reflexive (epistemicist) accessibility relation and [[.]] a function from atomic sentences to sets of 3D points of evaluation (each point being a subset of W × W × W). The bivalent valuation of sentences at 3D-points of evaluation $\langle w, v, u\rangle$ is defined as follows (P stands for an atomic sentence and ϕ and ψ are schematic variables):

$\lvert \mathrm{P} \rvert_{w,\, v,\, u} = 1$	iff	$\langle w, v, u\rangle \in [\![\mathrm{P}]\!]$
$\lvert \neg\phi \rvert_{w,\, v,\, u} = 1$	iff	$\lvert \phi \rvert_{w,\, v,\, u} = 0$
$\lvert \phi \wedge \psi \rvert_{w,\, v,\, u} = 1$	iff	$\lvert \phi \rvert_{w,\, v,\, u} = 1$ and $\lvert \psi \rvert_{w,\, v,\, u} = 1$
$\lvert \Box\phi \rvert_{w,\, v,\, u} = 1$	iff	for all $u^* \in \mathrm{W}$, $\lvert \phi \rvert_{w,\, v,\, u^*} = 1$
$\lvert A\phi \rvert_{w,\, v,\, u} = 1$	iff	$\lvert \phi \rvert_{w,\, v,\, v} = 1$
$\lvert \Delta\phi \rvert_{w,\, v,\, u} = 1$	iff	for all $u^* \mathrm{R} w$, $\lvert \phi \rvert_{u^*,\, v,\, w} = 1$

Again, ϕ is true *simpliciter* in w if $\lvert \phi \rvert_{w,\, w,\, w} = 1$. Logical consequence is defined as truth preservation in all proper points of evaluation (points of the form $\langle w, w, w\rangle$) in all models.

The first problem for the 3D semantics for definiteness comes from cases like "'Chancy' refers to the winner of the lottery" or "'tallness' refers to tallness". The solution that worked for the 4D model (insisting that 'refers' and 'Chancy' have diagonally true metasemantic character) comes at a price on the 3D model. The reason why Yli-Vakkuri (2016, p. 823) rejects the solution is that on his 3D model, the above solution implies that 'Chancy' and 'refers' are indexicals. On a 3D model, the diagonally true metasemantic character becomes a diagonally true character (i.e. a function f such that for all $w \in \mathrm{W}$, $\mathrm{f}(w)(w) = 1$). However, postulating a diagonally true character for expressions like '"τ" refers to τ' and 'Chancy is the winner of the lottery' would mean that 'Chancy' and 'refers' are indexicals. Suppose that we are using Yli-

Vakkuri's 3D model: postulating a diagonally true character for 'Chancy is the winner of the lottery' would mean that 'Chancy' refers to whichever ticket is the winner in the context world, e.g. 'Chancy' as evaluated at $\langle @, v, v \rangle$ would refer to the ticket winner at v. However, this would mean that 'Chancy' is an indexical. This is problematic because terms like 'Chancy' are not normally taken to be indexicals: they are rigid designators introduced by reference fixing descriptions.

These issues arise because of the limitations of the 3D model that are not present in the 4D model. In the 3D model, the second index (the context) is overloaded: it has to play the standard role of determining the intensions of indexical expressions as well as the role of determining the intensions of non-indexical expressions like 'Chancy'. This is the reason why expressions like 'Chancy' look like indexicals on the 3D model. However, on the 4D model these two roles are played by two separate indexes: the intension-determining role for indexicals is played by the context (the third index) and the character-determining role for expressions like 'Chancy' is played by the metasemantic circumstance (the second index). On a 4D model, neither 'true', nor 'refers, nor 'Chancy' are indexicals because their characters are constant: the proposal requires that they have diagonally true *metasemantic characters* (functions from worlds to characters) and not *characters* simpliciter (functions from worlds to contents). By moving to a 4D model, we are able to avoid the issues that a 3D model faces.

Another challenge to the 3D model comes from some examples of use of the actuality operator. The actuality operator is an indexical operator, which allows us to refer to the actual world. Yli-Vakkuri (2016, p. 824) presents an argument against epistemicist treatment of definiteness on his 3D model using a cleverly constructed non-indexical operator which is used to refer to the actual world.

Suppose that there is a non-indexical way of referring to the actual world. For instance, suppose that instead of using the standard indexical operator A, we can simply name the actual world 'Worldy' and refer to it simply using that proper name. We can then define a different actuality operator A^*; '$A^*\phi$' would roughly translate to 'In Worldy, ϕ'. Using his 3D model, Yli-Vakkuri (2016, p. 824) gives the following truth conditions to both the standard and the novel actuality operators:

$$|A\phi|_{w,\,v,\,u} = 1 \qquad \text{iff} \qquad |\phi|_{w,\,v,\,v} = 1$$

$$|A^*\phi|_{w,\,v,\,u} = 1 \qquad \text{iff} \qquad |\phi|_{w,\,w,\,w} = 1$$

The problem for the 3D-account of definiteness is that we seem to be in a position to know that in Worldy ϕ if and only if actually ϕ, i.e. our two ways of referring to the actual world should yield the same result. Therefore, '$A^*\phi \leftrightarrow A\phi$' should come out as definitely true. However, on the 3D semantics it does not come out as definitely true:

$$|\Delta(A^*\phi \leftrightarrow A\phi)|_{@,\,@,\,@} = 1 \qquad \text{iff} \qquad \text{for all } w^*\mathrm{R}@,\ |A^*\phi \leftrightarrow A\phi|_{w^*,\,@,\,@} = 1$$

$$\text{iff} \qquad \text{for all } w^*\mathrm{R}w,\ |A^*\phi|_{w^*,@,@} = |A\phi|_{w^*,@,@}$$

$$\text{iff} \qquad \text{for all } w^*\mathrm{R}w,\ |\phi|_{w^*,w^*,w^*} = |\phi|_{w^*,@,@}$$

As long as there is a close world w^* such that $|\phi|_{w^*,w^*,w^*} \neq |\phi|_{w^*,@,@}$ (which is guaranteed if there are to be differences between close possible worlds) the sentence will be false.

However, the issue does not arise for the 4D model. The reason again is that in a 4D model we have a complex metasemantic index consisting of a metasemantic context (first index) and a metasemantic circumstance. Firstly, the metasemantic context determines the metasemantic rule that the name 'Worldy' is to refer to whatever world is given by the metasemantic circumstance. Secondly, the metasemantic circumstance fixes that world that 'Worldy' is to refer to. Therefore, the truth conditions for the two actuality operators on the 4D model are given as follows:

$|A\phi|_{w,\,v,\,u,\,z} = 1$ iff $|\phi|_{w,\,v,\,u,\,u} = 1$

$|A^*\phi|_{w,\,v,\,u,\,z} = 1$ iff $|\phi|_{w,\,v,\,v,\,v} = 1$

However, now '$A^*\phi \leftrightarrow A\phi$' comes out as definitely true:

$|\Delta(A^*\phi \leftrightarrow A\phi)|_{@,\,@,\,@,\,@} = 1$ iff for all $v^*\mathrm{R}@$, $|A^*\phi \leftrightarrow A\phi|_{v^*,\,@,\,@,\,@} = 1$

iff for all $v^*\mathrm{R}w$, $|A^*\phi|_{v^*,\,@,\,@,\,@} = |A\phi|_{v^*,\,@,\,@,\,@}$

iff for all $v^*\mathrm{R}w$, $|\phi|_{v^*,\,@,\,@,\,@} = |\phi|_{v^*,\,@,\,@,\,@}$

As long as the second and third indexes (metasemantic circumstance and metasemantic context) are identical, 'A' and 'A^*' will be interchangeable.

This is the result given by the formal model, but it also makes sense given our intended interpretation. Suppose that the Twinglish speaker is with us in the actual world and is making the utterance 'In Worldy, ϕ'. What does 'Worldy' in their mouth refer to. Since in our scenario the Twinglish speaker lives in the actual world, 'Worldy' in Twinglish refers to the actual world. Like with 'Chancy', there is a clear metasemantic rule to determine the reference of 'Worldy'. Thus, in every close variant of the language 'Worldy' refers to the actual world: if we used a different close variant of the language, the reference of 'Worldy' would not change as all nearby language variants use the same metalinguistic rule to pick out the reference for 'Worldy'. Therefore, the results given by the formal model are in accordance with what we would expect from it.

7 Conclusion

I presented a multidimensional model inspired by the work of Stalnaker and Kaplan. Using that model I outlined an epistemicist proposal for the semantics of the definiteness operator. I showed that the presented account meets the desiderata and I argued that it is superior to the accounts presented in the literature.

Chapter 6

Vagueness of Truth

ABSTRACT

My aim in this chapter is to provide a solution to the T-Schema Puzzle presented by John Hawthorne (2006) as a problem for the epistemicist theory of vagueness. On a very natural interpretation, epistemicism seems to be committed to semantic plasticity of metalinguistic expressions such as 'truth'. However, it turns out that if 'truth' is semantically plastic then there are borderline instances of the Tarskian T-Schema. This would be a bad result for the epistemicist: we seem to know every instance of the T-Schema, so no instance can be borderline. I argue that if the truth predicate satisfies some intuitive principles, then 'truth' is not semantically plastic. Furthermore, I square the account of 'truth' with the account of definiteness developed in Chapter 5. I argue that my position is superior to Hawthorne's (2006), because it does not require accepting a problematic thesis that there are 'semantic joints in nature'.

WORD COUNT: 5,546

1 Introduction

In this chapter, I focus on one of the consequences of epistemicism for metalinguistic terms such as 'truth' or 'reference'.[94] As pointed out by John Hawthorne (2006), it turns out that when we apply the epistemicist theoretical machinery to metalinguistic vocabulary we get an interesting puzzle. The epistemicist theory seems to require that metalinguistic expressions such as 'truth' or 'reference' are vague, which would allow the epistemicist to explain why a sentence such as '"Michael is tall" is true' is borderline whenever the sentence 'Michael is tall' is borderline. It turns out that if 'truth' is vague, then we get borderline instances of the Tarskian T-Schema:

(T-Schema) $\ulcorner\phi\urcorner$ is true if and only if ϕ

However, it seems that we are in a position to know every instance of the T-Schema (putting aside issues that have to do with self-reference). Under the assumption that if a proposition is borderline then it is unknowable, no instance of the T-Schema can be borderline.

My aim in this chapter is to argue that 'truth' (and other metalinguistic terms) are not vague. In Section 2, I present the T-Schema Puzzle stated by Hawthorne (2006). In Section 3, I clarify the puzzle. In particular, I show that the puzzle arises only if we assume that the truth predicate plays a particular role in our language; I also show that if the truth predicate plays that role, then 'truth' is not vague. In Section 4, I discuss the implications of the puzzle for the epistemicist semantics and metasemantics.

[94] In the chapter I focus on 'truth', but everything that I say can be extended to expressions such as 'reference', 'extension', 'expresses' etc.

2 The T-Schema Puzzle

John Hawthorne (2006) presents an interesting sceptical argument. Hawthorne shows that given some intuitive epistemicist principles some instances of the Tarskian T-Schema come out as borderline and therefore unknowable. A caveat: I will present the puzzle using Hawthorne's (2006) setup. At some points, the setup differs from the strategy developed by me in the previous chapters. These differences will be explored in later parts of the chapter.

Epistemicism aims to explain the phenomena that are characteristic of vagueness, such as the propensity for generating borderline cases, without rejecting classical logic or bivalence. The epistemicist claim is that vagueness is associated with a particular kind of ignorance that gives rise to borderline cases. This particular kind of ignorance arises due to the semantic plasticity of expressions in our language. An expression is semantically plastic if and only if:

(a) small changes in the use of language produce small changes in the expression's intension AND

(b) we are insensitive to small changes in the expression's intension.

Given the safety condition on knowledge we can explain why semantic plasticity gives rise to a particular kind of ignorance.[95] If one expresses their belief using a sentence ϕ, then small shifts in the intension of ϕ could make ϕ false in some instances. Thus, in order for in order for ϕ to be definitely true, ϕ has to be true relative to all the interpretations found in nearby possible worlds.

Prima facie, it seems that 'truth' is a vague predicate. Suppose that (1) is borderline:

[95] See Williamson (2009) for a detailed discussion of the safety principle.

(1) Michael is tall.

Then, (2) is also borderline:

(2) 'Michael is tall' is true.

This seems obvious: if (due to vagueness) I am not in a position to know that Michael is tall, then I am also not in a position to know that 'Michael is tall' is true. On the standard epistemicist model, the following principle holds:

(Use Principle) If ϕ is borderline, then there is a constitutive expression used in ϕ that is semantically plastic.

In case of (2), the only likely candidate is 'true': 'is' is a standard copula and 'Michael is tall' is just a name of a sentence. Thus, it seems that 'true' is semantically plastic.

Hawthorne's (2006) argument is that if semantic plasticity of 'truth' is to explain the borderline status of sentences such as (2), then some instances of the Tarskian T-Schema come out as borderline. Consider an instance of the Tarskian T-Schema:

(3) 'Michael is tall' is true if and only if Michael is tall.

It seems obvious that we can know that 'Michael is tall' is true if and only if Michael is tall, even if 'Michael is tall' is borderline. That is because we seem to know every instance of the Tarskian T-Schema:

(T-Schema) $\ulcorner\phi\urcorner$ is true if and only if ϕ.

Suppose that 'truth' is semantically plastic: there is a close possible world w^* such that the truth predicate ('true*') in the language (L*) spoken at w^* has a different intension than our truth predicate ('true$^{@}$'). Suppose that the difference in intension between 'true$^{@}$' and 'true*' manifests itself in w^*, i.e. the extension of 'true$^{@}$' at w^* is distinct from the extension of 'true*' at w^*. (This is an assumption Hawthorne (2006) makes, but as I will show later it is

an innocent one). Therefore, let's suppose that 'Michael is tall' falls in the extension of 'true$^{@}$' at w^*, but not in the extension of 'true*' at w^*.

The problem is that if this is the case then an utterance of (3) made at w^* is false. The right-hand-side of (3) is true. That is guaranteed if we assume that our truth predicate (and all the truth predicates in the nearby worlds)[96] obeys the following principle:

> (Evaluation) For any ϕ and any world w, ϕ is true at w (ϕ is in the extension of 'true' at w) iff on the variant of the language spoken at w, ϕ is true at w.

Given the above principle and the proposition that 'Michael is tall' is in the extension of 'true$^{@}$' at w^*, we can conclude that 'Michael is tall' is true at w^*. However, the left-hand-side of (3) as used at w^* is false at w^*, because 'Michael is tall' is not in the extension of true* at w^*. Therefore, (3) is false at w^*. Since w^* is a nearby world, that makes (3) borderline.[97]

This is a bad result for the epistemicist. Firstly, it means that some instances of the T-Schema are unknowable.[98] Moreover, it means that our truth predicate could easily not obey the T-Schema, which is implausible given that the T-Schema plays an important theoretical role for us; even more strongly,

---

[96] We cannot assume that Evaluation holds for 'true' as used in any world, because 'true' in some distant variant of the language may have a completely different meaning (not corresponding in any way to our use of the truth predicate). Thanks to Ofra Magidor for this point.

[97] Note that the puzzle could be easily rephrased using other metalinguistic terms such as 'refers' or 'expresses', as such terms also come with analogues of the T-Schema such as: '$\ulcorner\tau\urcorner$ refers to τ' or '$\ulcorner\phi\urcorner$ expresses the proposition that ϕ'. So all that I say about 'truth' below could be said of other metalinguistic terms.

[98] The thesis that borderlineness is incompatible with knowledge is accepted by many theorists of vagueness, including canonical supervaluationists and epistemicists. However, this thesis has recently come under questioning by some philosophers (Barnett (2000), Barnes & Cameron (2011), Dorr (2003), Williams (2012)). Nevertheless, the epistemicist is committed to the thesis that borderlineness is incompatible with knowledge, so I will not question it here.

we seem to know that 'truth' obeys the T-Schema. These consequences are troubling. In what follows I will try to clarify the puzzle and outline potentials solutions.

3 Clarifying the puzzle

3.1 The role of truth

The key step in the puzzle is the move that allows us to evaluate a sentence uttered in a non-actual variant of the language using the actual truth predicate. This is justified if we assume that our truth predicate obeys Evaluation. However, Evaluation is not a trivial thesis about the role and behaviour of the truth predicate. If Evaluation holds, then we can use the truth predicate in our language to evaluate sentences belonging to other languages; according to Evaluation, what it is for '⌜ϕ⌝ is true' to be true at some world is for ϕ as used at that world to be true in that world. For instance, suppose that w is a French-speaking world (assume that there is one language per world for simplicity). We would want to be able to say things like '"la neige est blanche" is true in w' to mean that the sentence 'la neige est blanche', relative to how the speakers in w use it, is true in w. If the truth predicate obeys Evaluation, then our language has expressive power to evaluate sentences belonging to other languages.

There are good reasons to think that our truth predicate (and truth predicates in all nearby worlds) can be used to evaluate sentences belonging to other languages. We definitely talk as if we were able to evaluate sentences belonging to other languages in the manner described above. However, one may insist that the notion of truth that obeys Evaluation is not the only useful notion of truth. For instance, a reasonable alternative to Evaluation is Redundancy:

(Redundancy) For any ϕ, ϕ is true at w if and only if ϕ relative to how it is actually used is true at w.

On such a notion of truth, the truth predicate is not used to evaluate sentences belonging to other languages, but rather to evaluate sentences belonging to the actual variant of the language in counterfactual circumstances. Nevertheless, I will assume that we should proceed as if our truth predicate can be used to evaluate sentences belonging to other languages: this is something that we ordinarily do and is a useful feature of our language. Therefore, I will not pursue routes to solving the puzzle that deny our language's power to evaluate sentences that belong to other languages. However, when we get down to discussing the details of the semantics for 'truth', we will need to be careful with regards to how we represent this expressive power to evaluate sentences in other languages as not to end up with the problem pointed out by Hawthorne.

3.2 The strengthened puzzle

One may attempt to deal with the puzzle by denying various steps taken by Hawthorne in deriving the puzzle. In particular, one may try to tinker with the semantics for truth and definitions of definiteness; as shown in Chapter 5 there are multiple ways of defining definiteness and some are not so vulnerable to the puzzle. I will not consider all such options. Nevertheless, I believe that the puzzle can be strengthened to make a general point about the vagueness of truth.

In setting up the puzzle, Hawthorne assumes that if 'true' is semantically plastic, then there is a nearby world w^* such that 'true$^@$' (the actual truth predicate) and 'true*' (the truth predicate belonging to the variant spoken at

w*) have different extensions in w*. That assumption is need not be true.[99] For two predicates to have a different intension all it takes is for there to be *some* possible world where their intensions differ, it need not be any particular world. Therefore, the assumption that the extensions of 'true$^{@}$' and 'true*' differ in w* in particular is not an innocent one.

Nevertheless, even if we relax Hawthorne's assumption, we can still generate a version of the puzzle that leads to similarly problematic consequences. Suppose that there is some world w such that the extensions of 'true$^{@}$' and 'true*' are distinct at w. Then there is a sentence ϕ such that it is in the extension of 'true$^{@}$' at w but not in the extension of 'true*' at w. Suppose that p is the proposition expressed by ϕ relative to w. Then one of the below sentences is false at w:

(1) If $\ulcorner\phi\urcorner$ says that p, then $\ulcorner\phi\urcorner$ is true$^{@}$ if and only if p.

(2) If $\ulcorner\phi\urcorner$ says that p, then $\ulcorner\phi\urcorner$ is true* if and only if p.

But (1) and (2) are just instances of the basic disquotational principle for truth that the truth predicate obeys according to epistemicism (Williamson 1993, p. 188). Therefore, one of the truth predicates, 'true$^{@}$' or 'true*', violates the disquotational principle for truth at w.

The disquotational principle for 'truth' is one of the key commitments of epistemicism. For instance, Williamson's (1994b, p. 188) argument for bivalence crucially relies on the disquotational principle. Therefore, the epistemicist should treat the disquotational principle as one of the penumbral constraints on the meaning of 'true'.[100] That is to say that any candidate for the meaning of the word 'true' has to obey that principle: satisfying the disquotational principle is a feature of any predicate that plays the truth role

[99] Thanks to Ofra Magidor for pointing this out to me in a conversation.

[100] For an explication of penumbral constraints and connections see Fine (1975).

in any nearby variant of the language. If this is the case, then at least one of 'true$^{@}$' or 'true*' violates the penumbral constraint.

The result generalises: whenever we have two candidate truth predicates that yield different truth values when evaluating a sentence belonging to some language, one of the candidates fails to satisfy the disquotational principle. We have managed to generate a version of the puzzle that does not rely on Hawthorne's (2006) assumption that the differences in extension between our two predicates has to be manifested in any specific world. The outcome of this version of the puzzle is slightly different: we avoided a threat to the T-Schema, but there is a threat to the more general disquotational principles for truth that are endorsed by the epistemicist. Therefore, the semantic plasticity of 'truth' still poses a problem for the epistemicist.

An important caveat should be made here. A crucial feature of the argument is that there is no shift in the variant of language relative to which we evaluate the sentence using both of the truth predicates. For suppose that we evaluate a sentence like 'Michael is tall'. Suppose that it is true relative to the actual world, but false relative to the nearby world w*. Even if the two predicates 'true$^{@}$' and 'true*' differ with regards to how the sentence is evaluated, this does not pose a problem if we tacitly shift the variant of the language relative to which the sentence is to be evaluated. For in that case, 'Michael is tall' will express different propositions relative to these different variants of the language, which will not present a problem for the disquotational principles above.[101]

[101] Of course this does not mean that there are no worlds where 'true' does not obey the disquotational principles. This is obvious: there is nothing magical about the world 'true' and its meaning could be different (e.g. it could have the same meaning as our word 'red' is some world).

It might be objected that my use of the disquotational principle is inappropriate here. One might claim that the principle only applies to sentences that belong to the same language as the truth predicate used by the principle. This objection does not work, because we have assumed that the truth predicate in our language (and all nearby variants) can be used to evaluate sentences in other languages. So the disquotational principle should not be restricted.

In the next section, I will explore the consequences of the puzzle for the semantics for definiteness presented in Chapter 5 and the epistemicist metasemantics.

4 Consequences of the puzzle

4.1 Semantic consequences - truth and definiteness

Consider again the semantics for the definiteness operator presented in Chapter 5.[102] Suppose that our 4D model is a triple $\mathrm{M} = \langle \mathrm{W}, \mathrm{R}, [\![.]\!] \rangle$. As usual, W is a set of worlds, R is the accessibility relation, and $[\![.]\!]$ is the interpretation function. The valuation at 4D-points is defined as follows (P is an atomic sentence and ϕ, ψ are schematic variables):

$|\mathrm{P}|_{w,\, v,\, u,\, z} = 1$ iff $\langle w,\, v,\, u,\, z \rangle \in [\![\mathrm{P}]\!]$

$|\neg\phi|_{w,\, v,\, u,\, z} = 1$ iff $|\phi|_{w,\, v,\, u,\, z} = 0$

$|\phi \wedge \psi|_{w,\, v,\, u,\, z} = 1$ iff $|\phi|_{w,\, v,\, u,\, z} = 1$ and $|\psi|_{w,\, v,\, u,\, z} = 1$

$|\phi \leftrightarrow \psi|_{w,\, v,\, u,\, z} = 1$ iff $|\phi|_{w,\, v,\, u,\, z} = |\psi|_{w,\, v,\, u,\, z}$

[102] Quick recap: the account of definiteness works as a metasemantic monster that shifts the first index in the quadruple $\langle w,\, v,\, u,\, z \rangle$. This index represents the metasemantic rules used to assign metasemantic characters (functions from worlds to characters) to expressions. A sentence is definitely true on the account if it is true on all the assignments of metasemantic characters to expressions given by every nearby world.

$$|\Delta\phi|_{w,\,v,\,u,\,z} = 1 \qquad \text{iff} \qquad \text{for all } w^*\mathrm{R}w,\ |\phi|_{w^*,\,v,\,u,\,z} = 1$$

Additionally let's add a predicate T corresponding to our notion of truth. Let's begin with the proposal for the semantics of T that obeys Redundancy:

$$\text{(Redundant T)}\ |T\ulcorner\phi\urcorner|_{w,\,v,\,u,\,z} = 1 \qquad \text{iff} \qquad |\phi|_{w,\,v,\,u,\,z} = 1$$

Such a notion of truth simply strips off quotation marks off ϕ. It always evaluates the sentences relative to how they are used in the context of utterance (if $\langle w,\ v,\ u,\ z\rangle$ is the context in which $T\ulcorner\phi\urcorner$ is used, then T evaluates ϕ relative to the context $\langle w,\ v,\ u,\ z\rangle$). Redundant T has the nice feature that it automatically verifies all the instances of the biconditional $T\ulcorner\phi\urcorner \leftrightarrow \phi$. Thus, there is no threat of an instance of the T-Schema becoming borderline.

One thing to note about Redundant T is that it's semantically plastic. If we understand semantic plasticity as a difference in the function that is assigned to expressions by the metasemantic context (first index), then T comes out as semantically plastic.[103] However, the semantic plasticity of Redundant T does not give rise to the T-Schema Puzzle (or its strengthened version). Redundant T is only used to evaluate sentences belonging to the same variant of the language as the truth predicate itself. But as noted above, the puzzle only arises if we use 'truth' to evaluate sentences belonging to different languages.

However, this same feature that prevents Redundant T to be safe from the puzzle makes it deficient as an interpretation of the notion of truth that we are after. As argued in Section 3.1, an important role that the notion of truth plays is that it allows us to evaluate sentences belonging to different languages.

---

[103] The function assigned at metasemantic context is a function from triples $\langle v,\ u,\ z\rangle$, that correspond to the second, third and fourth index in the semantics, to truth values. Since there is a nearby world w^* such that $|T\ulcorner\phi\urcorner|_{w^*,@,@,@} \neq |T\ulcorner\phi\urcorner|_{@,@,@,@}$, T is semantically plastic.

Since Redundant T does not allow us to do this, it's lacking as the notion of truth that we are after.

Suppose then that we use a notion of truth that satisfies Evaluation. Then our truth predicate will evaluate sentences relative to how they are used in the circumstance of evaluation (last index). Then T would have the following semantics:

(Evaluative T) $|T\ulcorner\phi\urcorner|_{w,\,v,\,u,\,z} = 1$ iff $|\phi|_{z,\,z,\,z,\,z} = 1$

Such a notion of truth can be easily used to evaluate sentences uttered in other languages: a sentence ϕ is in the extension of T at a world z if and only if ϕ as used by the linguistic community in z is true at z. Therefore, Evaluative T fixes the issue that Redundant T suffered from. However, Evaluative T again poses problems for instances of the T-Schema. Suppose that w^* is a nearby world such that for some sentence ϕ, $|\phi|_{w^*,@,@,@} \neq |\phi|_{@,@,@,@}$. Then the biconditional $T\ulcorner\phi\urcorner \leftrightarrow \phi$ is not definitely true.[104]

Therefore, what we want something between Redundant T and Evaluative T: a notion of 'truth' that allows us to evaluate sentences belonging to other languages and does not give rise to issues with the T-Schema. As I can see it there are two ways of achieving this. Let's consider them in turn.

The first approach is to start with Redundant T as the default semantics for truth and supplement it with a way of directly referring to other worlds (variants of the language) that would specify the world relative to which we are to evaluate the sentence. Suppose that to evaluate a sentence relative to

[104] $|\Delta(T\ulcorner\phi\urcorner \leftrightarrow \phi)|_{@,@,@,@} = 1$ iff for all wR@, $|T\ulcorner\phi\urcorner \leftrightarrow \phi|_{w,@,@,@} = 1$
iff for all wR@, $|T\ulcorner\phi\urcorner|_{w,@,@,@} = |\phi|_{w,@,@,@}$
iff for all wR@, $|\phi|_{@,@,@,@} = |\phi|_{w,@,@,@}$
But we know that there is a nearby world w^* such that $|\phi|_{@,@,@,@} \neq |\phi|_{w^*,@,@,@}$. So $|\Delta(T\ulcorner\phi\urcorner \leftrightarrow \phi)|_{@,@,@,@} = 0$.

how it is used at w^*, we add a qualifier 'at w^*' to the use of T. The truth conditions for such a construction are given by:

(Redundant T+) $|T\ulcorner\phi\urcorner \text{ at } w^*|_{w,\, v,\, u,\, z} = 1$ iff $|\phi|_{w^*,\, w^*,\, w^*,\, w^*} = 1$

Adding the qualifier would allow us to evaluate sentences relative to how they're used at other worlds.[105] Moreover, if Redundant T would be a default option when no qualifier is added, then we would have no issues with borderline instances of the T-Schema.

The second option is to try to complicate the truth conditions for our truth predicate without introducing the qualifier referring explicitly to worlds. Suppose that we start with Evaluative T as the default semantics for T. To fix the semantics for T we should first determine what the problem with Evaluative T as an account of our truth predicate was. The problem with the indefiniteness of the T-Schema emerges because in the case when $T\ulcorner\phi\urcorner$ is evaluated at a point $\langle w, @, @, @\rangle$, the semantics for Evaluative T tell us that this is true when ϕ is true at $\langle @, @, @, @\rangle$. But this is not right: when we evaluate $T\ulcorner\phi\urcorner$ at $\langle w, @, @, @\rangle$ our intention is not to evaluate sentences belonging to the variant of the language spoken in @ (and this is what Evaluative T interprets it as). Consider an instance of the T-Schema: "'Michael is tall' is true iff Michael is tall". The reason why we take it to be definitely true is that no matter what truth conditions are assigned to the right-hand

[105] We can make the construction in a few ways. One way is to treat 'at w^*' as an operator that shifts the last index in the quadruple $\langle w, v, u, z\rangle$. Let's assume that T when it is combined with the qualifier 'at w^*' has the semantics of Evaluative T. The derivation would then work as follows:
$|T\ulcorner\phi\urcorner \text{ at } w^*|_{w,\, v,\, u,\, z} = 1$ iff $|T\ulcorner\phi\urcorner|_{w,\, v,\, u,\, w^*} = 1$ iff $|\phi|_{w^*,\, w^*,\, w^*,\, w^*} = 1$
This would mean that the predicate T is ambiguous between Redundant T and Evaluative T. When no qualifier is added, T gets the reading given by Redundant T. When the qualifier 'at w^*' is added, the reading changes to the semantics given by Evaluative T.

side of the biconditional ('Michael is tall'), the biconditional is guaranteed to be true because the sentence is about the language it is uttered in. However, Evaluative T makes it seem as if $T\ulcorner\phi\urcorner$ evaluated at $\langle w, @, @, @\rangle$ is about the variant of the language spoken at @. Therefore, we need to fix the semantics for T so that it reflects the intuitive idea that the relevant instances of the T-Schema are about the variant of the language they are uttered in.

What we need at the minimum is for the semantics of T to make it the case that for all nearby worlds w, $|T\ulcorner\phi\urcorner|_{w,@,@,@} = |\phi|_{w,@,@,@}$. We can make this happen by introducing the restriction that the metasemantic character of $T\ulcorner\phi\urcorner \leftrightarrow \phi$ is *diagonally true*. Recall that a metasemantic character of an expression is the function, determined by the metasemantic context (the first index), from triples $\langle w, v, u\rangle$ to truth values. The metasemantic character of an expression is diagonally true if it is true for any proper triple $\langle w, w, w\rangle$, i.e. a triple where each element of the triple is identical. Then the truth conditions for T are given by:

(Evaluative T+) $|T\ulcorner\phi\urcorner|_{w,\ v,\ u,\ z} = 1$ iff
(i) $v = u = z$ and $|\phi|_{\mathrm{w,\ z,\ z,\ z}} = 1$ OR
(ii) $(v \neq u \vee u \neq z \vee v \neq z)$ and $|\phi|_{\mathrm{z,\ z,\ z,\ z}} = 1$

Point (i) guarantees that when the second, third and fourth index are identical, Evaluative T+ will evaluate $T\ulcorner\phi\urcorner$ as if it would be about the variant of the language to which it belongs. In such a case, T works in the same way as the Redundant T. On the other hand, if the second, third, and fourth indexes are not all identical, Evaluative T+ will evaluate $T\ulcorner\phi\urcorner$ as if it were about the variant of the language in the circumstance of evaluation.

Evaluative T+ may seem to be an inelegant solution, but it allows T to perform two important tasks. Firstly, it allows us to evaluate sentences belonging to other languages. Secondly, it allows us to have a notion of truth

that does not pose problems with our definiteness operator. The notion of definiteness that was developed in Chapter 5 is based on the idea that we evaluate sentences belonging to other variants of the language as if they were uttered in the actual world. This is the interpretation of ϕ being evaluated at $\langle w, @, @, @\rangle$, where w is a nearby world. However, we would like it to be possible for the sentence belonging to the otherworldly variant of the language (and uttered in the actual world) that makes use of T to be about that otherworldly variant as opposed to the actual variant. Evaluative T+ allows us to achieve both of these desiderata.

Therefore, it seems that we can adjust our semantics for 'truth' so that it does not raise issues with borderline instances of the T-Schema and avoids Hawthorne's puzzle.

4.2 Metasemantic implications of the puzzle

The strengthened version of the puzzle shows that if 'truth' is used to evaluate sentences belonging to other languages (without a tacit shift in the variant of the language that we are evaluating), then the claim that 'truth' is semantically plastic is problematic. If there are two candidates for the notion of 'truth', then at least one of them fails to satisfy the disquotational principles that we know 'truth' obeys. Therefore, the candidates that don't satisfy the disquotational principles can be excluded as they don't meet the relevant criteria for the notion of 'truth': only one candidate is eligible.

We can use this idea to challenge an objection to epistemicist metasemantics posed by Rosanna Keefe (2000, Chapter 3). Epistemicism requires that every vague predicate draws a sharp boundary between things in its extension and things in its anti-extension. For example, for some number N, N is the smallest number of dollars that one can have and be rich; one dollar

less, and one is not rich. Thus, the boundary between things that fall in the extension of 'rich' and those that fall in the anti-extension of 'rich' is drawn at N: those that have N dollars or more are in the extension of 'rich' and those that do not, are not. But the objection goes, why should we care about the notion of *extension* rather than about a very similar notion of *extension**, which differs from *extension* only in that it is slightly more demanding: only things that have N+1 dollars are in the *extension** of 'rich'. Of course, saying that we should care about extension and not extension* is because all and only those things that 'rich' is true of are in the extension of 'rich', but are not in the extension* of it. The reply from the objector is to concede that it is not the case that all and only things that 'rich' is true of are in the extension* of 'rich', but there is a very similar notion to *truth*, namely *truth**, and all and only things that 'rich' is true* of fall in the extension* of 'rich'. Therefore, we have multiplied the notions, but have not answered the original question: why should we care about extension and truth rather than about extension* and truth*?

What the strengthened puzzle shows is that, once the variant of the language is fixed, there is only one eligible notion of 'truth'. The extension of 'true' does not float free: it supervenes on the extensions of other expressions in the language. The reason why we care about truth and not truth* is that truth satisfies the disquotational principles and truth* does not: in some cases a sentence $\ulcorner\phi\urcorner$ will say that p, but the biconditional '$\ulcorner\phi\urcorner$ is true* if and only if p' will be false. So we shouldn't care about truth*.[106]

[106] This challenge to epistemicism has to be distinguished from the standard challenge of how the epistemicist explains the fact that for a vague predicate like 'rich', the usage facts in the linguistic community determine a particular cut-off point. The challenge just considered is about the boundary of metalinguistic vocabulary like 'truth', this one is about the boundary of 'rich'. The latter challenge is dealt with in Chapter 2.

5 Conclusion

The T-Schema Puzzle gives us an interesting insight into the epistemicist treatment of metalinguistic terms like 'truth'. It turns out that 'truth' is not semantically plastic if it plays a role of allowing us to evaluate sentences or utterances belonging to other languages, as well as the disquotational role that epistemicism is committed to.

Part III – Puzzles

The third part of the thesis consists of two chapters and is aimed at answering some recent puzzles. Chapter 7 focuses on a puzzle to epistemicist explanation of ignorance due to vagueness, presented by Adam Sennet (2012). Chapter 8 is somewhat different from the rest of the thesis. There I focus on the case of metalinguistic comparatives, i.e. sentences like 'his problems are more financial than legal'. In the chapter, I make use of some theoretical resources available to the epistemicist to introduce a novel account of metalinguistic comparatives. I think that the chapter indicates that there are applications of the framework developed for the purposes of elucidating the epistemicist notion of semantic plasticity to other (seemingly unconnected) areas of philosophy. I take this to be a strength of the project.

Chapter 7

The Omniscient Speaker Puzzle

ABSTRACT

The epistemicist theory of vagueness aims to explain ignorance due to vagueness by semantic plasticity: the disposition of intensions of vague expressions to shift with usage across close possible worlds. This explanation is challenged by the Omniscient Speaker Puzzle (Sennet 2012). Suppose that an omniscient speaker, Barney, who knows all the facts about usage and how these determine the intension of expressions cooks up a scheme to stabilise the intension of a normally semantically plastic term like 'rich'. It seems that 'rich' would display all the phenomena associated with vagueness without being semantically plastic, thus making the epistemicist explanation of vagueness insufficient. In this chapter, I present for a few choice points for epistemicism that arise as a result of the puzzle.

WORD COUNT: 5,659

1 Introduction

The epistemicist theory of vagueness associates the phenomenon of a borderline case with a particular kind of ignorance. Specifically, the epistemicist explains the ignorance due to vagueness by appealing to the semantic plasticity of vague expressions. A natural interpretation of the epistemicist theory is that semantic

plasticity is a necessary condition for vagueness: if ϕ is vague, then ϕ is semantically plastic. However, an interesting challenge may be posed for the epistemicist: there are cases where a term exhibits all the phenomena associated with vagueness and yet is not semantically plastic (Sennet 2012).

In this chapter, I will show that the epistemicist has a way of meeting such challenges. In Section 2, I will briefly outline Williamsonian epistemicism. In Section 3, I present the challenge to epistemicism coming from the Omniscient Speaker Puzzle (Sennet 2012). In Section 4, I show that in the light of the puzzle, there are several choice points for the epistemicist. Ultimately, the puzzle does not turn out to be a devastating objection to epistemicism, but forces the epistemicist to come up with explanations that push the theory forward.

2 Epistemicism

The prince is clearly rich and the pauper is clearly not, but there are borderline cases where we are not in a position to know whether someone counts as being rich or not. Suppose that Dan has $40,000 and is a borderline case of richness. According to the epistemicist, there is a boundary between the rich and the non-rich: there is a smallest amount of dollars one needs to have to count as being rich. The key question for the epistemicist theory is to explain why we are not in a position to know in borderline cases, e.g. why we are not in a position to know whether Dan is rich. The epistemicist answer is Semantic Plasticity:

> (Semantic Plasticity) An expression ϕ is semantically plastic if and only if, unbeknownst to us, ϕ's intension could easily be different.

The epistemicist is in a position to explain ignorance in borderline cases using Semantic Plasticity and the safety condition on knowledge:

(Safety) A belief is safe if and only if it could not easily be the case that one has the belief and it is false.

The epistemicist takes safety to be a necessary condition on knowledge. The combination of Safety and Semantic Plasticity allows the epistemicist to explain ignorance in borderline cases. Suppose that Dan just falls on the side of being rich. The reason why a belief that Dan is rich does not constitute knowledge is that it is not safe. Since 'rich' is semantically plastic, it could easily denote a slightly different property, richness*, which is slightly more demanding so that Dan does not count as being rich*. Therefore, it could easily be the case that we would say something false by expressing our belief that Dan is rich (say, by uttering 'Dan is rich'). So ignorance due to vagueness is explained by the failure of Safety due to semantic plasticity in borderline cases.

Thus, there is an interesting interplay between knowledge, Semantic Plasticity, and Safety. Safety is a necessary condition for knowledge. So failure of Safety is sufficient for ignorance. However, it also seems that the epistemicist wants to explain ignorance in borderline cases by semantic plasticity. Ignorance in borderline cases is explained by the failure of safety due to semantic plasticity. A natural interpretation of epistemicism commits the theory to the following Central Thesis:

(Central Thesis) If ϕ is vague, then ϕ is semantically plastic.

In the next section I will present a puzzle due to Adam Sennet (2012) that challenges the Central Thesis.

3 The Omniscient Speaker Puzzle

Consider a linguistic community where the cut-off point for 'rich' is $40,000. Additionally, imagine that one of the speakers in the community, Barney, is

an omniscient speaker. Barney is able to detect all uses of the term 'rich' within the linguistic community. He also knows how usage of 'rich' determines its intension for the community. Barney is dedicated to keeping the cut-off point for 'rich' fixed at $40,000. Whenever someone makes an utterance that would shift the intension of 'rich', Barney makes sure that a counter-utterance is made to perfectly offset the shift in the intension and keep it fixed at $40,000. Consider a member of the linguistic community, Jessica, who believes that (#) is true:

(#) The cut-off point for 'rich' is $40,000.

Sennet (2012) argues that there are a few propositions that we can make about the case.

(1) *'Rich' is not semantically plastic in the community.* If Barney is committed to keeping the intension of 'rich' fixed at $40,000, then it does not seem like the intension could easily shift. The intension of 'rich' is stable across all nearby worlds.

(2) *'Rich' is vague in the community.* Sennet (2012, p. 278-279) argues that 'rich' exhibits all the phenomena that vagueness is associated with. Since Jessica and other members of the community other than Barney don't know about Barney's scheme to keep the intension of 'rich' stable, they will treat 'rich' as any other vague term, e.g. they will exhibit the same kind of uncertainty in borderline cases for 'rich' as in case of any other vague term etc.

(3) *Jessica does not know that the cut-off point for 'rich' is at $40,000.* Even if Jessica has a true belief about the cut-off point for 'rich', she is not in a position to know the location of the cut-off point any more than we do.

This constitutes a puzzle for the epistemicist. Firstly, (1) and (2) are jointly inconsistent with the Central Thesis: if epistemicism relies on the Central

Thesis, this is a problem for epistemicism. Secondly, Jessica is not in a position to know that (#) is true. However, 'rich' in Jessica's community is not semantically plastic, so it cannot be semantic plasticity that explains Jessica's ignorance. Since epistemicism postulates that vague expressions have sharp boundaries, it seems that the burden is on the epistemicist to explain Jessica's ignorance of the sharp cut-off for 'rich'.

4 Choice points for epistemicism

Since (1) and (2) are jointly inconsistent with the Central Thesis, there are three strategies that the epistemicist might adopt to try to solve the puzzle: deny (1), deny (2), or deny that epistemicism is committed to the Central Thesis.[107] In the next section, I will briefly say why denying (1) is not a fruitful strategy. The plan for the later sections will be to consider the choice points for epistemicism when it comes to dealing with the puzzle.

4.1 A weak strategy: insisting on the semantic plasticity of 'rich'

Let's consider the question whether 'rich' is semantically plastic in Barney's and Jessica's community. 'Rich' is clearly not semantically plastic for Barney

[107] It's possible to argue that epistemicism is not strictly speaking committed to the Central Thesis. The epistemicist claims that borderlineness is explained by semantic plasticity, but it is not committed to the more specific thesis that the borderlineness of ϕ is explained by the semantic plasticity of ϕ (and not some other sentence). For instance, consider the sentence '"Michael is tall" is true'. Suppose that it is borderline. It is entirely possible that the borderline status of '"Michael is tall" is true' is explained by the semantic plasticity of another sentence: 'Michael is tall'. So the epistemicist is not committed to the Central Thesis in full generality. However, solving the problem is not that simple. Whereas in the case of '"Michael is tall" is true' it is not the semantic plasticity of the sentence itself, but rather the semantic plasticity of another sentence, that explains the borderlineness, in case of (#) there does not seem to be any other sentence that could explain the (potential) borderline status of (#). So pursuing this strategy of solving the puzzle does little for the epistemicist.

since he is an omniscient speaker: the intension of 'rich' could not be easily different unbeknownst to him. The question is whether it is semantically plastic for Jessica and other members of the community. 'Rich' is semantically plastic for them only if there is a close possible world where, unbeknownst to them, the intension of 'rich' is slightly different. Clearly, there are worlds where the intension of 'rich' is slightly different; for instance, Barney could decide to stop his stabilising scheme or decide to keep it fixed at some different value. The question is whether such worlds would count as relevantly *close*.

One can imagine an epistemicist strategy along the following lines. The epistemicist could resist the problem simply by claiming that worlds where Barney decides to fix the reference of 'rich' at a different value (or where he abandons his scheme) should count as relevantly close. Such worlds are definitely epistemically possible for Jessica: what she knows does not rule them out as she's unaware of Barney's scheme and not an omniscient speaker herself. So it's not clear that 'rich' is not semantically plastic after all.

However, the effectiveness of such a strategy is limited, since we are not able to clearly state all the relevant conditions that would tell us which worlds count as relevantly close.[108] Ultimately, the worlds that count as relevantly close for the purposes of safety are those that are relevant for knowledge, i.e. are similar enough that a mistake in such a world would count as an easy mistake relative to the actual world. Since it's not clear which worlds count as close, what the epistemicist achieves at best is a position in which it's not clear whether the epistemicist has a problem. At worst, this insisting that 'rich' is semantically plastic despite the stabilising scheme sounds like grasping at straws. The epistemicist should do better.

[108] See Williamson (2000).

4.2 The Non-Ambitious Route

In Chapter 4 I outlined two versions of epistemicism: Ambitious and Non-Ambitious. The Ambitious Epistemicist is interested in explaining definiteness and borderlineness: this was the route I have taken in Chapters 4, 5 and 6. On the other hand, the Non-Ambitious Epistemicist does not treat borderlineness as a joint-carving notion. The Non-Ambitious Epistemicist is interested in explaining ignorance. The Omniscient Speaker Puzzle does not pose a problem for the Non-Ambitious Epistemicist, because it should not be a problem for the epistemicist to explain Jessica's ignorance of the truth of (#).

The epistemicist explains ignorance by pointing to failures of safety. Suppose that, as the epistemicist claims, vague predicates like 'rich' have sharp cut-off points. One natural question that arises is why we cannot know where these sharp cut-off points are. The standard Non-Ambitious Epistemicist response is that semantic plasticity is partly responsible for our ignorance: the sharp boundaries of vague predicates shift in ways inscrutable to us, so our beliefs regarding the location of these boundaries would be unsafe and thus could not constitute knowledge. However, even if 'rich' is not semantically plastic in Jessica's community, we may still explain her ignorance by appealing to other factors.[109]

The standard epistemicist explanation of ignorance in borderline cases arises straightforwardly from the account of inexact knowledge presented by

[109] Of course, the epistemicist may also adopt a more confrontational strategy and insist that there is nothing to explain. After all, ignorance, and not knowledge, is our natural state. Unless the critic presents an argument for why we should expect Jessica to possess the relevant knowledge, there is nothing for the epistemicist to explain. Safety is merely a necessary condition for knowledge: even if it turns out that a belief is safe it does not mean it constitutes knowledge. However, this is a less dialectically effective strategy for the epistemicist, because the critics expect epistemicists to explain why the sharp cut-offs for vague expressions postulated by epistemicists are unknowable.

Williamson (1992). Suppose that there are 10,000 leaves on a tree that I am looking at. I know that there are more than 100 and less than 1 million leaves on the tree, but I don't know the exact number of leaves on the tree. Furthermore, suppose that I luckily estimate and come to believe correctly that there are exactly 10,000 leaves on the tree. On a safety conception of knowledge that the epistemicist embraces one shows that this belief does not constitute knowledge by showing that there is a nearby world where the belief is false. There are two standard ways of showing this: the first is to show that the facts with which the belief is concerned could themselves be different (which would make the belief false); the second is to show that the method used to obtain the luckily true belief could easily produce a false one. The first way involves keeping the belief state fixed and finding a nearby world where the facts are different. The second way attempts to find a nearby world where the facts in the world with which the belief is concerned are perhaps not different, but the belief (produced by the same method as in the actual world) held in that world is false.

In attempting to show that my true belief that there are 10,000 leaves on the tree is unsafe, we could firstly note that the number of leaves on a tree is modally plastic: it could easily be different. Therefore, we might look at a nearby world where the number of leaves is slightly different, but I still believe that there are 10,000 leaves on the tree; this world would be a witness for the thesis that my belief is unsafe (using the first way outlined above). The second standard way of showing that my belief is unsafe is to look at nearby worlds where using the same method that I employ actually I reach a slightly different conclusion about the number of leaves on the tree. Since my method of estimating the number of leaves (looking at the tree without any aids) is not very reliable, there is a nearby world where I, using the same method, estimate the number of leaves to be something else than it is in that world, e.g. I could

easily estimate that the number of leaves is 10,001, but the number of leaves would still be 10,000.[110]

Our knowledge of the meanings of vague expressions in our language is an example of inexact knowledge, e.g. we are in a position to know that 'tall' picks out people who are over 200 cm in height and does not pick out those who are below 160 cm in height, but we are unable to say exactly where the boundary for 'tall' is. Therefore, in Jessica's case, we may attempt to employ the standard ways of explaining ignorance in cases of inexact knowledge. Since the boundary for 'rich' is kept fixed by the omniscient speaker in the community, we assume that there are no nearby worlds where the boundary for 'rich' is different than in the actual world. However, for Jessica's belief about the boundary of 'rich' to be safe, she would need to have a reliable method for arriving at her belief. If she merely correctly guesses where the boundary for 'rich' is, her belief would not be safe as there is a nearby world where she guesses incorrectly. Therefore, the Non-Ambitious Epistemicist can explain Jessica's ignorance in the puzzling case.[111] Since, the Non-Ambitious Epistemicist claims there is nothing more for them to explain, it seems that the Omniscient Speaker Puzzle is not a problem for the Non-Ambitious Epistemicist.

However, the work developed in Chapters 4-6 pursued the Ambitious route. The Omniscient Speaker Puzzle poses a problem for this approach. The account of definiteness presented in Chapter 5 relies on evaluating sentences

---

[110] The first way is used more often to demonstrate the unsafe status of beliefs. However, the second way is also employed, especially to deal with problematic cases like explaining our ignorance of necessary truths. Why cannot I know whether Goldbach's Conjecture is true? Answer: there is no proof of the conjecture and my method of arriving at any belief regarding the conjecture (guessing) could easily lead me to make an error even if I actually guess correctly.

[111] It should be noted that knowledge of the boundary for 'rich' is possible in Jessica's case: for instance, if Barney's scheme is made public then it's quite likely that the speakers in the community may come to know what the boundary for 'rich' is. I come back to this point below.

relative to interpretations taken from nearby possible worlds.[112] If the Omniscient Speaker is stabilising the reference of 'rich' in the way outlined above, then (#) is definitely true on the account presented in Chapter 5. As I see, there are two options for the epistemicist to have a chance in dealing with this problem. First, we might try to tweak the account presented in Chapter 5 as to avoid the problem. Second, we could take the error-theoretic route of trying to explain away the intuition that 'rich' is vague despite not being semantically plastic within the community in question. The next two sections are devoted to these two strategies.

4.3 The Normality Route

The epistemicist explanation of ignorance is crucially based on the safety condition on knowledge. Roughly, a belief is safe if and only if there is no nearby possible world where the belief is false. The safety condition on knowledge is a well-established (though not universally accepted of course) principle in epistemology. However, in the recent literature[113] there has been a shift from safety-based explanations of ignorance to normality-based ones. Whereas safety-based explanations aim to evaluate beliefs at nearby (similar) cases, normality-based explanations look at normal cases. Normality-based explanations aim to account for Gettier-cases just like safety-based explanations. For instance, the reason why I don't have knowledge that it's 12 o'clock while looking at a broken clock that shows 12 o'clock is that the process of belief formation is not normal (Loets 2022, p. 159-160). Normality-based

[112] Technically, the account relies on taking metasemantic characters (functions from worlds to characters) from nearby worlds. The distinction is not important here.

[113] For instance, Ball (2013), Beddor & Pavese (2020), Dutant (2016), Goodman (2013), Goodman & Salow (2018), Greco (2014, 2016), Leplin (2007), Littlejohn & Dutant (2020), Peet & Pitcovski (2018), Smith (2010, 2016, 2018a, 2018b), Stalnaker (2005, 2015).

explanations of ignorance could be potentially used to show that even though (#) is safely true (true in all the nearby cases), it is not true in all the normal cases.

Suppose that we change slightly the definition of semantic plasticity. Instead of defining a term as semantically plastic if and only if it's metasemantic character shifts across close nearby worlds, we can define it as a term whose metasemantic character is *disposed* to shift. Call the property to be disposed to shift in this way 'dispositional semantic plasticity'. The Omniscient Speaker Puzzle does not pose an objection to the claim that 'rich' is dispositionally semantically plastic: I think it's pretty clear that the term 'rich' is dispositionally semantically plastic in Barney's community. Barney's reference-fixing scheme is a fink: even though the intension of 'rich' is disposed to shift, through Barney's strange scheme the intension is kept stable. This case is quite similar to the standard examples of finkish dispositions discussed in the literature.[114] In a standard case, if x has a finkish disposition to ϕ in case ψ, then whereas in normal circumstances x would ϕ if it were the case that ψ, in the strange local circumstances it would not ϕ if it were the case that ψ. For instance, imagine that there is a porcelain vase that is fragile, but that would not break in any close counterfactual case because the local deity has decided to protect the vase at all costs. If a rock were thrown at the vase the deity would solidify the air around the vase; if the vase fell, the deity would make the floor soft etc. Similarly, in Barney's case, normally the intension of 'rich' has the disposition to shift with usage; however, in the strange local circumstances of Barney's community, the intension of 'rich' would not easily shift, because of Barney's stabilising scheme. Of course the strange finkish circumstances don't mean that we shouldn't identify fragility with propensity

[114] See Lewis (1997).

to break or associate ignorance in borderline cases with the effects of semantic plasticity. In other words, even if ignorance of the reference of 'rich' cannot be explained by semantic plasticity in Barney's community, it does not mean that the vagueness of 'rich' in our normal circumstances is not explained by semantic plasticity.

Sennet (2012, p. 280-281) argues that the analysis of Barney's case as an instance of a finkish disposition does not help the epistemicist. According to Sennet, we can adjust the case so that it involves nothing finkish. Suppose that Barney is not an omniscient speaker and has no scheme to stabilise the intension of 'rich'. However, the intension of 'rich' is stable nevertheless, because whenever someone makes an utterance that would shift the intension, a counteracting utterance just happens to be made by another member of the community thus keeping the intension stable.

The problem with Sennet's reply is that whereas it is possible that in a single possible world w the intension of 'rich' is luckily kept fixed across some span of time, we cannot suppose that in any world close to w the intension does not shift. To paraphrase Kripke (1972, p. 44), possible worlds are stipulated not discovered: there is a possible world w^* very similar to w where the intension of 'rich' is slightly different – for all intents and purposes that world will count as close. In the absence of finks like Barney's scheme, there is no reason to suspect that in all the worlds close to w the intension of 'rich' will be identical.

What does establishing that 'rich' in Barney's and Jessica's community is dispositionally semantically plastic get us? We still need to show how to run the standard explanations of ignorance of the cut-off point and ignorance in borderline cases using the notion of dispositional semantic plasticity. The scope of this chapter does not allow for detailed discussion of how the epistemicist account could be revised to give the notions of dispositional semantic plasticity

and normality-based explanations of ignorance the centre stage. I will sketch two directions such a project might go in.

The overall picture of this Dispositional Epistemicism is this. Vagueness is associated with dispositional semantic plasticity: ϕ is vague if and only if it is dispositionally semantically plastic. If ϕ is dispositionally semantically plastic, then the metasemantic character of ϕ will shift across the normal cases. The reason why dispositional semantic plasticity produces ignorance is that we combine it with a normality-based explanation of ignorance (replacing the safety-based one). Roughly, on such a conception, a belief constitutes knowledge only if it is true in all the normal cases. Dispositional semantic plasticity will produce errors in normal cases in the same way as semantic plasticity produces errors in nearby cases, which means that we will be able to explain ignorance using dispositional semantic plasticity.

On the account developed in Chapter 5, the definiteness operator took the metasemantic character from all the nearby worlds w and evaluated sentences relative to these characters (while keeping other indexes fixed). As mentioned above, I see two ways to adjust that account to the normality-based explanation of ignorance. Firstly, we might go for the *normal counterpart approach*. Suppose that instead of taking metasemantic characters from all the worlds that are close to the actual world, we take them from all the worlds that are close to a *normal counterpart* of the actual world. We could amend the account to include a function N, which for each possible world would find the closest normal counterpart of that world. In case the actual world (or a world relative to which we are evaluating definiteness claims) is a normal one, this account would work exactly the same way as the standard account. However, when the world in question is not a normal one, the function N shifts the centre world (world relative to which other worlds count as close) to a normal counterpart of the world in question. This would solve the Omniscient

Speaker Puzzle: Jessica's world is not a normal one, since there is a strange reference-fixing scheme in place. If we evaluate (#) relative to a normal counterpart, we will find it to be borderline as there is no reference-fixing scheme in normal worlds.

The second direction we might take to amend the account of definiteness, would be to amend the accessibility relation on worlds. Instead of understanding accessibility as an appropriate kind of similarity between worlds, we could employ a notion of accessibility understood as normality. A world w is accessible from v iff w is a normal counterpart of v. An account of definiteness using such an accessibility relation would be identical to the standard account, apart from the interpretation of the accessibility relation. On such a version of epistemicism, (#) would also count as borderline, because there are normal counterparts (one's on which the reference-stabilising mechanism is not in place), on which (#) is false. Thus, the puzzle could be solved by such an account.

These are just sketches of normality-friendly versions of epistemicism. Lots more work would be required to make them alternatives to the standard epistemicist account that could merit and allow for assessment. I am not convinced by the normality account; I prefer the road more travelled by. However, I believe that there are good reasons to think that a normality-based version of epistemicism could handle problems like the Omniscient Speaker Puzzle. Furthermore, I think the above sketches of the accounts are good starting points for future research.

In the next section, I will outline the error-theoretic route to solving the puzzle. My aim will be to deny that 'rich', as is used by the Omniscient Speaker's community, is vague. I will also attempt to explain why it nevertheless may seem to be vague without being so.

4.4 The Error-theoretic Route

Sennet (2012) argues that in the puzzle case, 'rich' exhibits all the phenomena associated with vagueness. That is inconsistent with the standard epistemicist picture: semantic plasticity is one of the manifestations of vagueness; if Sennet is correct that 'rich' is not semantically plastic in Jessica's and Barney's community, then a key component of vagueness is missing by the epistemicist standards. However, we can assume that all members of the community (apart from Barney) would treat (#) as if it were borderline – it would seem borderline, and therefore vague, to them. How to explain their treatment of 'rich' as vague?

The first thing to note is that the inference from 'ϕ is not borderline' to 'definitely ϕ is not borderline' is not valid. If it were the case, then the members of the community could not be mistaken about the borderline status of (#). However, this inference is not valid. For instance, D. W. Hart (1992) argues that the term 'heap of sand', probably most often used in the literature as a paradigm case of vagueness, is not in fact vague (and hence has no borderline cases): the smallest stable heap structure consists of 4 grains of sand (three grains on the bottom and one on the top), so 'heap' has a knowable sharp boundary at 4 grains. Of course we don't have to accept Hart's arguments regarding 'heap of sand'. Nevertheless, his position is plausible; it would not be ridiculous to suppose that 'heap of sand' is not in fact a vague term even though, pre-theoretically, it seems to be so. Therefore, it is not absurd to propose that we would be wrong about some definite cases being borderline. More importantly, there are also more general reasons to think that the inference from 'ϕ is not borderline' to 'definitely ϕ is not borderline', i.e. an inference from $(\neg\nabla\phi)$ to $(\Delta\neg\nabla\phi)$, is not valid. For if it were valid, there would

be no higher order vagueness.[115] Similarly, the inference from $\nabla\phi$ to $\Delta\nabla\phi$ is invalid for the same reasons. We could be mistaken about what's borderline and what's definitely true.

The second thing to note is that the rest of the Omniscient Speaker's community is unaware of the reference-stabilising scheme. They treat the term 'rich' as any other vague term. If the scheme were made publicly known, then there would be no puzzle. Suppose that the Omniscient Speaker announces his scheme to the community. Then, other members of the community would be in a position to know that (#) is true, because learning what the reference of 'rich' is from an Omniscient Speaker who stabilises its meaning is a good (safe) method of obtaining knowledge about its reference. Therefore, the epistemicist may plausibly claim that in the puzzle case the term 'rich' is not in fact vague: if it were vague there would be no reliable method of learning what the reference is.

Nevertheless, it seems that there is still a task for the epistemicist to complete. The epistemicist has to explain why members of the community would think that the term is vague despite the fact that its reference is stable (so should not count as vague by epistemicist standards). I think that despite the fact that we are not perfectly reliable when it comes to assessing whether a term is vague, our intuitions about what's vague are right most of the time. That is to say the inference from 'ϕ seems vague' to 'ϕ is vague' is a pretty reliable heuristic.[116] Even if Hart (1992) is right that 'heap' is not in fact vague, which would show that this heuristic is not perfectly reliable, there are not many such examples: we are correct about what's vague most of the time.

---

[115] The statement $(\neg\nabla\phi) \rightarrow (\Delta\neg\nabla\phi)$ entails that there is no higher order vagueness, because it entails the (4) axiom: $\Delta\phi \rightarrow \Delta\Delta\phi$ (if we assume that Δ obeys the standard K and T axioms for modal logic).

[116] See Williamson (2020) on recent philosophical applications of reliable heuristics.

Therefore, we should expect the community to expect that 'rich' is vague, even though (by epistemicist terms) it is not: they are simply relying on a pretty reliable heuristic.

Note that this explanation does not only work for the puzzle at hand, but also in cases where the we are mistaken about some term being vague for some other reason. For instance, in the case of Hart's (1992) argument against the vagueness of 'heap', it seems that the reason why 'heap' is not vague is that there is a (previously uncovered) joint-carving referent for the term 'heap'. So we could also (potentially) be wrong about a term being vague, because there is a previously uncovered joint-carving (natural) referent for some term we thought to be vague. Similar mistakes could occur in case of when (unbeknownst to the majority of a population) a term is introduced by a precise stipulation. Suppose that w is a world like ours apart from the fact that the term 'bald' was first introduced as a term referring to those with 1000 hairs or less. However, the linguistic community at large is not aware of this initial stipulation: they use 'bald' in the same way as we do. Nevertheless, arguably, 'bald' as used at w has a stable sharp boundary at 1000 hairs, unbeknownst to the members of the community at w. We need to explain why to the community at w the term 'bald' would seem vague: the explanation provided above can serve that purpose. Therefore, it seems that we need an explanation for why we would be mistaken about a term being vague that's independent of the puzzle.

5 Conclusion

The Omniscient Speaker Puzzle can produce a variety of responses from the epistemicist. There are a few choice points for the epistemicist. Firstly, the epistemicist has to decide whether to take the Ambitious or the Non-Ambitious

route. The puzzle does not present a problem for the Non-Ambitious Epistemicist, but their Ambitious cousin does have some explaining to do. Secondly, the epistemicist has to decide whether to stick to the standard account based on safety explanations of ignorance or whether to pursue a new path of normality-based explanations. Two ways of running the normality-based project were sketched; however, more work needs to be done to properly assess such proposals. Lastly, if the epistemicist sticks to the standard explanation of definiteness, they should take the error-theoretic route: argue that 'rich' is not vague because of the reference stabilising scheme and explain why it seems to be vague for the members of the community. It seems that such an explanation is available: it seems that the community relies on an inference from 'ϕ seems vague' to 'ϕ is vague'; though this inference is not always truth preserving, it is pretty reliable. Such an explanation of why we can be mistaken about what's vague is needed for independent reasons: it could be that a term is not vague despite appearances because of some unknown reference magnet or uncovered initial stipulation. Overall, the Omniscient Speaker Puzzle forces the epistemicist to make some important decisions about the development of the theory and to push it forward.

Chapter 8

Metalinguistic Comparatives and Semantic Plasticity

ABSTRACT

Some comparative statements such as 'Adam's problems are more financial than legal' evade capture by the traditional accounts of comparatives: they involve terms that are not gradable and have a specific metalinguistic flavour to them. The nature of metalinguistic comparatives (MCs) is often summed up in the phrase that they are in a sense about what's better to say. In this chapter, I provide a novel account of metalinguistic comparatives. I employ the notion of similarity between interpretations of language (inspired by Timothy Williamson's notion of semantic plasticity) and argue that using this notion we can analyse metalinguistic comparatives. Roughly, an MC 'X is more F than G' is true if and only if the interpretations according to which 'X is F' is true and 'X is G' false are more similar to the actual interpretation of our language than interpretations on which the former is false and the latter is true.

WORD COUNT: 6,993

1 Introduction

Not all comparatives are created equal. As the recent literature suggests there is a class of expressions, metalinguistic comparatives (MCs), that differ from ordinary comparatives in important ways. For example, metalinguistic comparatives, as opposed to ordinary comparatives, can involve terms that are not gradable as in (1):

(1) Adam's problems are more financial than legal.

Furthermore, MCs can involve terms of varying syntactic categories.[117] One way of cashing out the difference between ordinary comparatives and metalinguistic comparatives is that MCs seem to be (at least partly) about language. Ordinary comparative expressions, such as 'John is taller than Mike' or 'playing the violin is more difficult than playing the viola', are wholly about the world (e.g. John and Mike's height). On the other hand, MCs like (1) are partly about the world (the nature of Adam's problems), but also about language (i.e. whether Adam's problems should be classified as 'financial' or 'legal'). This metalinguistic flavour of MCs is usually summarised by the phrase that MCs are in a sense about 'what's better to say'. For instance, to say that Adam's problems are more financial than legal is to say that it is in some sense better to say that Adam's problems are financial than to say that his problems are legal. The task is to figure in what sense is it 'better to say' one thing rather than another in the context of using metalinguistic comparatives and to provide semantics for MCs.

---

[117] Rudolph & Kocurek (2020, p. 2-3) provide the following list of examples:
- Line is more [DP a syntactician] than [DP a semanticist].
- A chimp is more [NP ape] than [NP monkey].
- Harriet more [VP tumbled into a solution] than [VP sought it].
- The dog is sitting more [PP on your head] than [PP in your lap].

My aim in this chapter is to provide a novel interpretation of metalinguistic comparatives. Firstly, I will present my account of MCs based on semantic similarity and motivated by the framework used by Williamsonian epistemicists to analyse semantic plasticity. Secondly, I will argue that other proposals found in the literature are lacking.

2 Desiderata

There are several key desiderata for the account of metalinguistic comparatives.

A. If we are to interpret an MC 'X is more F than G' as saying that it's in a sense better to say 'X is F' than to say 'X is G', we must clarify in what relevant sense it would be better say one thing as opposed to another.

B. Consider the following MC:

(2) He is more a logician than a philosopher

The easiest cases of MCs are ones where there is a difference in truth value between 'he is a logician' and 'he is a philosopher'. If he is a logician but not a philosopher, it's easy to explain why 'he is more a logician than a philosopher' would be true. (Similarly, it's easy to explain why (2) would be false if he were a philosopher but not a logician). The difficult case arises when 'he is a philosopher' and 'he is a logician' both have the same truth value. Any account of MCs must account for the single-value cases.

C. MCs often involve non-gradable terms, e.g. 'logician'. Consequently, when analysing an MC like (2) we cannot simply compare the degrees to which a person possesses the property of being a logician and the property of being a philosopher. Nevertheless, we need to construct an

ordering on which 'he is a logician' scores higher than 'he is a philosopher' to provide our account of MCs.

In the next section I will outline my view and I will show that it meets all the above desiderata.

3 The Similarity View

3.1 Semantic similarity

Williamsonian epistemicism argues that what is characteristic about vagueness is a special kind of ignorance. When I look at Michael, I may be ignorant as to whether Michael is tall. My ignorance could be due to the fact that I am looking at Michael from a distance and I am unable to determine his height exactly. Alternatively, I could be ignorant whether Michael is tall, because Michael is a borderline case of 'tall'. The epistemicist argues that ignorance in borderline cases is a result of certain features of concepts and expressions that would be used to express the belief that Michael is tall. More specifically, our expressions and concepts are shifty: their meanings could easily be different. This feature of our concepts and expressions is called semantic plasticity.

An expression[118] α is semantically plastic if and only if unbeknownst to us the intension of α shifts across the space of close possible worlds. The basic idea behind the epistemicist theory is that we could use the words we use differently, e.g. we could use the word 'red' in the way that we actually use the word 'rich'. Moreover, since we could easily use language differently than we do, the intensions assigned to expressions in our language could easily be

[118] The same treatment is applied to concepts. I will focus on expressions from now on for the sake of simplicity.

different. For instance, if we would use the term 'tall' more conservatively than we do, then the cut-off point for 'tall' would be higher than it actually is.

The picture painted by the epistemicist is the following. Across the space of possible worlds, the interpretations (assignments of intensions to expressions in our language) shift with usage. Suppose that the actual intension of 'tall' is identical to the intension of 'over 180 cm in height'. Moreover, suppose that w^* is a world where the intension of 'tall' is identical to our intension of 'over 190 cm in height'. It's natural to suppose that in the space of worlds the boundary of 'tall' does not suddenly jump from 180 cm to 190 cm. Plausibly, there is a series of worlds between the actual world and w^*, the boundary of 'tall' shifting slightly from one world to the next (e.g. as a result of small shifts in usage). This is the Argument from Parity discussed in Chapter 2. The ordering of the series is determined by how similar the intension of 'tall' in a given world is to the actual intension of 'tall'; the higher the boundary for 'tall', the further away (relative to the actual world) the world where 'tall' has that boundary. The basic idea is to use this ordering to provide the account of metalinguistic comparatives.

However, before we generate the desired ordering, we have to explicate the notion of similarity that we will rely on. In order to generate the desired ordering on interpretations, I will rely on a similarity relation between interpretations. The task of spelling out the details of how to understand the similarity between interpretations can be tricky. For instance, suppose that to generate the similarity relation on interpretations, we would rely on an overall similarity relation between worlds (e.g. as assumed by Lewis (1973) in his analysis of counterfactuals). Let us call our needed similarity relation on interpretations 'semantic similarity'. Let w_1, w_2… be worlds, @ the centred world and i_1, i_2… interpretations of our language. Let w_1 be a world such that (i) i_1 is the interpretation used at w_1 and (ii) i_1 is not used in any world w^* that

is overall more similar to @ than w_1 is. Similarly, let w_2 be a world such that (i) i_2 is the interpretation used at w_2 and (ii) i_2 is not used in any world w^* that is overall more similar to @ than w_2 is. We can define 'semantic similarity' ($\leq$) as follows: $i_1 \leq i_2$ if and only if w_1 is at least as close (at least as overall similar) to @ as w_2 is.[119]

There is an obvious problem with such an interpretation. Suppose that an all-powerful demon has a switch that would make us use the term 'tall' to have the same meaning as the actual term 'red'. In the actual world the switch is not on, but there is an overall very similar (very close) world w^* where the switch is on: on the interpretation at w^*, 'tall' applies to red things. On the account based on overall similarity, the interpretation at w^* counts as very similar to the actual interpretation. Nevertheless, intuitively the interpretation at w^* is not close to the actual interpretation (in the sense of 'close' required for our analysis).

Considering that the notion of overall similarity relation between worlds raises some problems, we should try to go via different route. Let's start with the intuitive notion of similarity between interpretations. There is an intuitive sense in which interpretations can be similar to one another. Suppose that N cm is the actual boundary for 'tall'. The interpretation according to which the boundary is at N+1 cm is, *ceteris paribus*, more similar to the actual interpretation than an interpretation according to which the boundary is at N+2 cm simply because the difference between N and N+1 is smaller than the difference between N and N+2. (Alternatively, the shortest man who's height

[119] Intuitively, the closer a world w is to the actual world, the smaller the 'distance' from the actual world to w. Hence, if w_1 is more similar to @ than w_2 is, then $i_1 \leq i_2$ (the 'distance' from the actual interpretation to i_1 is smaller than the 'distance' from the actual interpretation to i_2).

is over N+1 cm is closer in height to the shortest man over N cm in height than the shortest man over N+2 cm in height is.)

I think the intuitive sense of similarity between interpretations is pretty clear when it comes to predicates such as 'tall'. One might worry that this notion is more difficult to track when it comes to expressions such as 'is a logician' or 'is a philosopher' which don't denote easily quantifiable properties. However, there is also an intuitive sense of similarity of interpretations when it comes to such expressions; for instance, there are interpretations that allow for a more liberal use of terms such as 'is a logician' and ones that are stricter. Interpretations that are only slightly more liberal are more similar to the actual interpretations than interpretations that are very liberal. Therefore, the intuitive understanding of the similarity relation between interpretations also applies to such cases. Nevertheless, there is a question about how to formalise this intuitive notion of similarity.

The Argument from Parity, discussed in Chapter 2 and rehashed above, makes use of the intuitive notion of similarity between interpretations. One worry we might have is that the argument only relies on some partial notion of similarity. It's one thing to say that relative to a particular predicate like 'tall' one interpretation is more similar to the actual interpretation than another; it is another thing to postulate that there is a total (with respect to the entire language) relation of similarity on interpretations.

However, we can borrow the notion of similarity from the epistemicist analysis of definiteness (such as the account developed in Chapter 5). The epistemicist account of definiteness requires a total relation of similarity on interpretations. The account of definiteness in Chapter 5 uses an accessibility relation R in defining the definiteness operator. At the very least, what we get from that account is a total notion of similarity on interpretations such that all the interpretations in the accessible worlds are more similar to the actual

interpretation than the interpretations in the worlds that are not accessible from the actual world (@). We can also extend the notion of similarity to the interpretations in worlds non-accessible to @. Say that all the interpretations in the worlds accessible from @ are similar to the actual interpretation to degree 1. Then, we can say that all the interpretations in worlds that are not accessible from @, but accessible from a world that is accessible from @, are similar to the actual interpretation to degree 2. We can follow that procedure to define greater degrees of similarity of interpretations. This gets us a notion of similarity on interpretation: any interpretation similar to degree N is more similar to the actual interpretation than an interpretation similar to degree N+1. Therefore, it seems we can get a total notion of similarity on interpretations from the account of definiteness developed in Chapter 5.

Of course, this would get us a basic notion of similarity on interpretations on which we don't differentiate between interpretations that are assigned the same degree of similarity. Suppose that the actual boundary for 'tall' is 180 cm and that our margin for error is 5 cm: interpretations on which the boundary for 'tall' is less than 175 cm or more than 185 cm don't count as close (worlds where such interpretations hold are not accessible). On the basic notion of similarity I just outlined, an interpretation i_1 on which the boundary is 181 cm has the same degree of similarity 1 as the interpretation i_2 on which the boundary is 184 cm. The more reiterations of the margin for error are required to reach an interpretation, the greater it's degree of similarity. Since, both i_1 and i_2 are within the margin of error from the actual world, so they get assigned the degree of similarity 1. Nevertheless, intuitively the former is more similar to the actual interpretation than the latter.

However, once we have such a basic notion of similarity, there is no reason to suppose that we would not be able to refine it to allow us to differentiate between interpretations that are assigned the same degree of similarity. In the

example above, I assumed that the margin for error is 5 cm. However, suppose that we change the margin for error to 3 cm. Then, i_1 would still be within the margin of error, but i_2 would not (it would be assigned the degree of similarity 2). So by changing the margin for error, it seems that we can differentiate between i_1 and i_2. The former is more similar to the actual interpretation than the latter, because decreasing the margin of error made the latter one to fall outside of the margin. There remains a question as to whether we can stipulate that we can get the notion of similarity to be fine-grained enough that for any two interpretations, either one is more similar to the actual interpretation or the other one is. This question will be explored at the end of the next section. Nevertheless, it seems that we can borrow the notion of semantic similarity from the epistemicist account of definiteness.

3.2 The account

In what follows we will need the following ordering based on semantic similarity:

> (Semantic Similarity, $\leq$) For any interpretations i, j, i$\leq$j if and only if i is at least as semantically similar to the actual interpretation as j is.

What's the relationship between the semantic similarity relation and metalinguistic comparatives? The basic idea is the following. Consider again the sentence:

(3) Adam's problems are more financial than legal.

Saying that his problems are more financial then legal suggests the classification of his problems as 'financial' meets some higher threshold than classifying them as 'legal'. As discussed above, the easiest case to illustrate this is when his problems are financial but not legal: it is true that his problems

are financial, and it is false that they are legal so clearly classifying his problems as financial meets a higher threshold (truth) than classifying them as legal.

However, suppose that Adam's problems are both financial and legal. When it comes to classifying problems as 'financial' and 'legal', we could set the standards higher or lower. For instance, we could set standards very low so that anyone with unpaid debts over $10 counts as having financial problems. On the other hand, we could set the standards very high so that only people who are millions in debt count as having financial problems. The basic idea is that Adam's problems are more financial than legal if and only if Adam's problems would meet a higher standard when classifying them as financial than when classifying them as legal. Assuming that Adam's problems are both financial and legal, if we steadily increased the standards for both 'financial problem' and 'legal problem', Adam's problems would first cease to fall under the latter category. Using the notion of *semantic similarity*, the interpretation that is most semantically similar to the actual interpretation and where the truth values of 'Adam's problems are financial' and 'Adam's problems are legal' are different is one where the former is true and the latter one false (and not the other way round).

Let's consider the generic MC 'X is more F than G' in four cases: (i) X is F and X is not G, (ii) X is not F and X is G, (iii) X is both F and G, and (iv) X is neither F nor G. The first two cases are easiest. If 'X is F' is true but 'X is G' is false on the actual interpretation, then the MC is true because on the actual interpretation 'X is F' is true and 'X is G' is false and (tautologously) there is (a) no closer (more similar) interpretation on which 'X is F' would be false and 'X is G' would be true and (b) no closer interpretation on which 'X is F' and 'X is G' would both be false. For analogous reasons, the MC 'X is more F than G' is false if 'X is F' is false on the actual interpretation and 'X is G' is true.

Now let's consider the difficult single-value cases (iii) and (iv). If 'X is F' and 'X is G' are both true on the actual interpretation, then the MC is true if the closest interpretation where 'X is F' is true and 'X is G' is false is closer than (a) the closest interpretation where 'X is F' is false and 'X is G' true and (b) the closest interpretation where 'X is F' and 'X is G' are both false. That is to say that 'X is F' is more *robustly true* than 'X is G' is: our language would have to change more to make 'X is F' false than to make 'X is G' false. On the other hand if both 'X is F' or 'X is G' are both false, then the MC is true if 'X is F' is less robustly false: our language would have to change less to make 'X is F' true than to make 'X is G' true.

Let's try to formalise the account. Let L be our language containing names (a, b, c), predicates (F^n, G^n, H^n) and the comparative sentential connectives $\prec$ and $\approx$ expressing the metalinguistic comparatives 'more than' and 'as much as' respectively.[120] Let the well-formed formulas of the language be (A and B are metalinguistic sentential variables):

A: $F^n(a_1, \ldots, a_n) \mid \neg A \mid A\&B \mid A \vee B \mid A \rightarrow B \mid A \prec B \mid A \approx B$

In our semantics, formulas will be evaluated relative to an ordering $\leq$ on interpretations and interpretation-world pairs. Our model is a quintuple $M = \langle W, D, I, \leq, [\![.]\!] \rangle$ such that W is the set of worlds, D is the domain of objects, I a set of interpretations, $\leq$ is the ordering on interpretations as outlined above, and $[\![.]\!]$ is an interpretation function from atomic sentences of L to sets of interpretation-world pairs $\langle i, w \rangle$. For all $i \in I$:

- $i(a_1)$: $W \rightarrow D$ for each name a_1
- $i(F^n)$: $W \rightarrow \wp(D^n)$ for each predicate F^n (where $\wp$ stands for the powerset)
- $\langle i, w \rangle \in [\![F^n(a_1, a_2 \ldots)]\!]$ iff $\langle i(a_1)(w), i(a_2)(w) \ldots \rangle \in c(F^n)(w)$

[120] I follow the setup of Rudolph & Kocurek (2020) here.

Then let the bivalent valuation of sentences be defined as follows:[121]

$|A|_{\leq,i,\ w} = 1$ iff $\langle i,\ w\rangle \in [\![A]\!]$

$|\neg A|_{\leq,i,\ w} = 1$ iff $|A|_{\leq,i,\ w} = 0$

$|A \wedge B|_{\leq,i,\ w} = 1$ iff $|A|_{\leq,i,\ w} = 1$ and $|B|_{\leq,i,\ w} = 1$

$|A \succ B|_{\leq,\ i,\ w} = 1$ iff $\exists i^*$:

(i) $|A|_{\leq,i^*,\ w} = 1$ and $|B|_{\leq,i^*,\ w} = 0$ AND

(ii) $\neg\exists i^{**}$: $|B|_{\leq,i^{**},\ w} = 1$ and $|A|_{\leq,i^{**},\ w} = 0$ such that $i^{**} \leq i^*$ AND

(iii) $\neg\exists j \exists j^*$: $|A|_{\leq,j,\ w} = |B|_{\leq,j,\ w} = 1$ and $|A|_{\leq,j^*,\ w} = |B|_{\leq,j^*,\ w} = 0$ such that $j \leq i^*$ and $j^* \leq i^*$.

$|A \approx B|_{\leq,\ i,\ w} = 1$ iff

(i) $|A|_{\leq,i,\ w} = |B|_{\leq,i,\ w}$ AND

(ii) $\forall i^*$: If ($|A|_{\leq,i^*,\ w} \neq |A|_{\leq,i,\ w}$ or $|B|_{\leq,i^*,\ w} \neq |A|_{\leq,i,\ w}$) then

$\exists j$: $|A|_{\leq,j,\ w} = |B|_{\leq,j,\ w}$ and $|A|_{\leq,j,\ w} \neq |A|_{\leq,i,\ w}$ such that $j < i^*$

Let's go through all three points (i)-(iii) of the truth conditions for $\prec$ to clarify the account. Condition (i) guarantees that A is consistent (if it's inconsistent it will not be true on any interpretation and world). Moreover, it guarantees that there is an interpretation on which A is true and B is false: if A is to be more robustly true than B, then there has to be such an interpretation. What we want for the truth conditions is to guarantee that if A is true and B false on the actual interpretation then 'A$\succ$B' is to be true; furthermore, if the value of A and B is the same, what we want is for A$\succ$B' to be true if the first change in truth values of either A or B that we encounter, as we move down the ordering of interpretations by semantic similarity, is one on which A is true and B false. This is what (ii) achieve. Condition (ii) guarantees that the closest interpretation where A is true and B is false is more similar to the actual

[121] Sometimes, for ease of exposition, I use the strict version of the ordering on interpretations. The strict version of the relation can be defined from the non-strict version in a standard way: $x < y$ iff ($x \leq y$ and $\neg(y \leq x)$).

interpretation than any interpretation where A is false and B is true. Condition (iii) guarantees that as we move to further away interpretations the first time that we encounter a change in the truth values is not one on which the truth value of both A and B changes. Suppose that A and B are both true: what we want to avoid is a situation in which our truth conditions make 'A≻B' true even if the first change we encounter (as we go down the ordering of interpretation) is one on which both A and B are false; it seems that in such a case we should validate 'A≈B'. Condition (iii) eliminates such unwanted cases.

Similar analysis applies to the connective ≈ intended to be interpreted as 'as much as'. The basic idea with both ≺ and ≈ is that as we move from the actual interpretation to less similar interpretation, the first time there is a change in the truth value on either side of the comparative, it is a change in the right direction. For instance, if both A and B are true, the truth conditions for ≺ guarantee that if the comparative 'A≻B' is true if the first change of truth value to either A or B (as we move from the actual interpretation to less similar ones) is one on which A is true and B is false. Similarly, the truth conditions for ≈ guarantee that the first change of truth value to either A or B is one where the truth value in changes in both.

The account meets all the outlined desiderata. Firstly, there is a clear sense of what it means for an MC 'X is more F than G' to express the notion that it's better to say 'X is F' than 'X is G'. It's better to say 'X is F' than to say 'X is G' if 'X is F' is more robustly true (or less robustly false). Consequently, the account provides an ordering on interpretations which underlies the account of metalinguistic comparatives in case of terms denoting non-gradable properties. Lastly, the account can account for the single-value case when 'X is F' and 'X is G' are both true (or both false) and yet 'X is more F than G' is true. Thus, all the desiderata are met.

There is an interesting question that emerges out of the literature (Rudolph & Kocurek 2020, p. 7) on metalinguistic comparatives: do MCs validate total comparability:

(Total Comparability). One of the following sentences is true:

(i) X is more F than G

(ii) X is as F as G

(iii) X is more G than F

It's unclear whether total comparability is desirable. It seems that there are cases where we would not want to commit to total comparability. Consider the following cases (Kocurek & Rudolph 2020, p. 7):

(4) Al is more wise than tall.

(5) Al is more tall than wise.

(6) Al is as much wise as tall.

It's likely that there are instances where we wouldn't want to be committed to any of these claims. Whether the account above validates total comparability depends on the assumptions we make about the semantic similarity relation. In particular, it depends on whether the semantic similarity relation obeys the uniqueness constraint:

(Uniqueness) For any two distinct interpretations i and j, either $i<j$ or $j>i$.

If Uniqueness holds, then the account validates total comparability. Suppose that 'X is F' and 'X is G' are both true according to the actual interpretation. If Uniqueness holds then there is a unique closest interpretation i* where at least one of the truth values of 'X is F' and 'X is G' changes: (i) if 'X is F' is false and 'X is G' is true on i* then 'X is more G than F' is true on the actual interpretation, (ii) if 'X is F' is true and 'X is G' is false on i* then 'X is more

F than G' is true on the actual interpretation, and (iii) if 'X is F' and 'X is G' are both false on i* then 'X is as F as G' is true on the actual interpretation.

On the other hand, if Uniqueness does not hold then Total Comparability is invalidated. Again, suppose that 'X is F' and 'X is G' are both true according to the actual interpretation. If we don't assume uniqueness then, there is no guarantee that there is a unique closest interpretation on which either of the truth value of 'X is F' or 'X is G' changes. Suppose that there are two equally close interpretations i and j such that 'X is F' is true and 'X is G' is false on i but the latter is false and the former true on j. Then neither 'X is more F than G' nor 'X is more G than F' nor 'X is as F as G' will be true.

On the account I presented there are no reasons to think that Uniqueness holds. Suppose that the boundary for 'tall' on the actual interpretation is N cm. It seems that non-actual interpretations on which the boundary for 'tall' is set at N+1 cm and N-1 cm respectively will be equally close to the actual interpretation. Therefore, most likely total comparability should be rejected.

It's perhaps desirable to point out that the account presented above is not meant to be a complete story on metalinguistic comparatives. The account presents a way for translating the similarity relation on interpretations into the semantics of metalinguistic comparatives. This is (merely) foundational work. Several possible refinements come to mind. Firstly, we might want to define the metalinguistic comparative not as applying to full sentences, but rather as applying directly to predicates in these sentences. Secondly, we might want to make the similarity relation on interpretations something that is context dependent. Take a sentence like 'Adam's problems are more financial than legal'. In some contexts, we might want to put weight on different aspects of Adam's problems; it might be the case that, on some aspects his problems are more financial than legal and on other aspects they are more legal than financial. Such refinements to the account are possible, but they are beyond

the scope of this chapter, the aim of which is to present the account in its basic form.

In the next section I will outline rival accounts and I will argue that they fail to meet our desiderata or are vulnerable to counterexamples.

4 Existing frameworks

In the recent literature, there is a number of proposals on how to analyse metalinguistic comparatives. These include analyses carried out using notions of:

(A) Comparing conventions (interpretations)

(B) Appropriateness

(C) Precision

(D) Degrees of belief

Rudolph & Kocurek (2020) provide an excellent outline and critique of (B), (C), and (D). Themselves, they develop the account favouring (A). I agree with their critique of the three accounts, so I will present it very briefly at the end of the chapter. The next section will be focused on discussing the (A)-type account developed by Rudolph & Kocurek (2020).

4.1 Comparing conventions

From the proposals developed in the literature thus far, the account presented by Rudolph & Kocurek (2020) is most similar to my own. Let an *interpretation* be an assignment of intensions to terms in a given language.[122] Their starting point is the idea that sentences in our language may be evaluated using a non-actual interpretation of the language. On their account of MCs when using a

[122] Rudolph & Kocurek (2020) use the term 'convention' instead of 'interpretation'.

metalinguistic comparative, a speaker is tacitly expressing a commitment to an interpretation (perhaps one that is not actually used by the linguistic community) relative to which the sentence is to be evaluated. Roughly, 'Adam's problems are more financial than legal' expresses the speaker's commitment to the interpretation according to which Adam's problems count as financial and not legal. In short, an MC 'X is more F than G' expresses the speaker's stronger commitment to interpretations according to which X counts as an 'F' as opposed to ones according to which X does not count as an 'F' but counts as a 'G'.

There are two problems with this account. Firstly, the it suffers from a technical problem: the truth-conditions for metalinguistic comparatives provided by Rudolph & Kocurek (2020) are vulnerable to counterexamples in single-value cases. Secondly, it's not clear whether there is an adequate sense of 'speaker commitment' that could make the account work. Let's begin with the technical problem.

The framework provided by Rudolph & Kocurek (2020) is very similar to my own. Whereas the foundation of my account is the semantic similarity relation, Rudolph & Kocurek (2020) rely on the ordering on interpretations (or interpretation-world pairs) generated by the speaker's commitments: for two interpretations i and j, $i \sqsubseteq j$ if and only if the speaker is committed to j at least as she is committed to i.[123] However, they give the following truth conditions to the metalinguistic comparative (Rudolph & Kocurek 2020, p. 13):

(A1) $|A \succ B|_{\sqsubseteq, i, w} = 1$ iff $\exists i^*$: $|A|_{\sqsubseteq, i', w} = 1$ and $\forall i^{**} \sqsupseteq i^*$: $|B|_{\sqsubseteq, i^{**}, w} = 0$

Later they refine the account as not to validate total comparability (Rudolph & Kocurek 2020, p. 16):

[123] Rudolph & Kocurek (2020) relativise the ordering of interpretations to worlds. Here the account is simplified as we never consider cases where the world of evaluation changes.

(A2) $|A \succ B|_{\sqsubseteq,\, i,\, w} = 1$ iff

(i) $\exists i^*$: $|A|_{\sqsubseteq,\, i^*,\, w} = 1$ AND

(ii) $\forall j : |B|_{\sqsubseteq,j,\, w} = 1 \Rightarrow \exists j^* \sqsupseteq j : |A|_{\sqsubseteq,j^*,\, w} = 1$ and

$\forall j^{**} \sqsupseteq j^* : |B|_{\sqsubseteq,\, j^{**},\, w} = 0$

The technical problem with both (A1) and (A2) is that they don't deal with the single-value case when both A and B are true.[124] Suppose that we substitute A with 'Gödel was a logician' and B with 'Gödel was a philosopher'. Consider the following MC:

(7) Gödel was more a logician than Gödel was a philosopher.

Gödel was both a logician and a philosopher (certainly made contributions to both). So the speaker ought to be committed to the best interpretations being ones on which both 'Gödel was a logician' and 'Gödel was a philosopher' are true. However, given the importance of Gödel's contribution to logic and the proportion of his work that was devoted to logic as opposed to philosophy it would be correct to say that he was more a logician than a philosopher. Neither (A1) nor (A2) can validate (7): the speaker cannot be committed to (7) and to the best interpretation being one on which both 'Gödel is a logician' and 'Gödel is a philosopher' are true.

Secondly, the account suffers from an interpretative problem. It is not clear how to understand the speaker commitment to one interpretation being 'better than' another interpretation. Consider the disagreement between Alice and Bob:

Alice: Tom is more tall than medium height.

Bob: Tom is more medium height than tall.

[124] What's interesting is that Rudolph & Kocurek (2020) account handles the single-value case when both A and B are false.

According to the analysis by Rudolph & Kocurek (2020) what Alice is saying is true if there is an interpretation i' according to which Tom counts as 'tall' and there is no interpretation i" better than i' according to which Tom counts as 'medium height'. On the other hand, what Tom is saying is true if there is an interpretation i* according to which Tom counts as 'medium height' and there is no interpretation i** better than i* according to which Tom counts as 'tall'. What decides the standards according to which one interpretation counts as better than another (or what makes the speaker to be committed to one interpretation as opposed to another)? It cannot be the subjective judgement or preference of the speaker as it would mean that speakers have too much power in determining the truth conditions of their utterances. If Bob is committed to the interpretation i* according to which Tom counts as 'medium height' more than to any interpretation according to which Tom counts as 'tall', then 'Tom is more medium height than tall' would be true in Bob's mouth even if Tom was the tallest man in the world. Members of a linguistic community, generally, cannot freely set the standards for evaluation of their utterances as long as they are speaking a shared language. (This holds even a speaker prefers an alternative way of speaking.)

However, if it is not subjective judgement or preference of the speaker that determines the ordering of interpretations, it is not clear what the standard is supposed to be. The standard assumption in discussion of vague boundaries, such as the boundary between 'tall' and 'medium height', is that no boundary is better than any other as such terms are not joint-carving in any sense. So an interpretation that places Tom on one side of the boundary is no better (in any standard understanding of 'better') than an interpretation which places him on the other side of the boundary; there seems to be no reason for the speaker to be committed to one boundary as opposed to another. Therefore,

we need a framework which would provide a clear interpretation for the ordering of interpretations.

4.2 Other accounts

This section covers alternatives to the account of metalinguistic comparatives presented in Section 3 and to the account given by Rudolph & Kocurek (2020). Most of the issues presented here were already raised by Rudolph & Kocurek (2020). The section outlines three accounts and presents an overgeneration problem as well as an undergeneration problem for each account.

4.2.1 Appropriateness

According to an (B)-type account[125], when a speaker utters an MC like 'Al is more wise than clever' what is effectively being said is that saying that Al is wise is more appropriate or desirable than saying that Al is clever. The account faces two problems.

Overgeneration Problem. One issue with such an account is that there are many instances where saying that X is F is more appropriate or desirable than saying that X is G, but saying 'X is more F than G' is incorrect (Rudolph & Kocurek 2020). For instance, if at someone's wake Herman says 'Sorry your mother croaked', it would be correct to say (8) but not (9) (Morzycki 2011, p. 47):

(8) It's more appropriate/desirable to say 'she passed' than 'she croaked'.

(9) #She more passed than croaked.

Undergeneration Problem. There are cases when an MC 'X is more F than G' is correct even though it is not more appropriate or preferable to say 'X is F' than to say 'X is G'. For instance, in the context of discussing what to put in

[125] See Giannakidou & Stavrou (2009) and Giannakidou & Yoon (2011) for a defence of such an account.

an obituary of an over-confident climber who died as a result of his bravado, it would be correct to say (10) but not (11) (Rudolph & Kocurek, p. 5):

(10) He was more rash than brave.

(11) #It's more appropriate/desirable to say 'he was rash' than 'he was brave'.

4.2.2 Precision

According to a (C)-type account, 'Al is more wise than clever' is true if and only if it is more precise to say that 'Al is wise' than to say that 'Al is clever'. *Being more precise* is analysed in terms of degrees: the degree to which Al possesses the property of being wise is greater than the degree to which he possesses the property of being clever.[126]

Overgeneration Problem. The issue with a precision-based account is that there are instances when it is more precise to say that X is F than to say that X is G, but saying that X is more F than G is incorrect. Suppose that Herman is 180 cm in height. Then uttering (12) is more precise than uttering (13), but uttering (14) sounds infelicitous (Rudolph & Kocurek 2020, p. 6):

(12) Herman is 182 cm in height.

(13) Herman is 185 cm in height.

(14) #Herman is more 182 cm in height than 185 cm in height.

Undergeneration Problem. There are also instances when uttering an MC 'X is more F than G' is correct, even though it is not more precise to say that 'X is F' that it is to say that 'X is G'. Let's return again to Kurt Gödel's example. Gödel was both a logician and a philosopher. It would be 100% precise to call Gödel a logician and 100% precise to call him a philosopher (so it would be equally precise to say 'Gödel is a logician' as 'Gödel is a philosopher').

[126] See Morzycki (2011) for a defence of this account.

However, it would still be correct to utter (15) (e.g. given Gödel's volume of work on logic and philosophy):

(15) Gödel was more a logician than a philosopher.

4.2.3 Degrees of belief

A (D)-type account analyses MCs in terms of speaker's credences (degrees of belief).[127] An MC like 'Al is more wise than clever' is true if and only if the speaker's credence in the proposition that 'Al is wise' is higher than her credence in 'Al is clever'.

Overgeneration Problem. One problem with this account is that the speaker's higher credence in *X is F* as opposed to *X is G* could have nothing to do with X being more F than G, but rather with the speaker's evidence. Suppose that John sees Amy picking up her daughter from school. John does not know Amy well; he does not know that Amy is an accountant. So his credence in (16) is greater than his credence in (17):

(16) Amy is a parent.

(17) Amy is an accountant.

Nevertheless, (18) is infelicitous (as uttered by John or by anyone else):

(18) #Amy is more a parent than an accountant.

Undergeneration Problem. Another issue with the credence-based account is that there are instances where an MC 'X is more F than G' is felicitous even if X is not more likely to be F than to be G. Suppose that Al has a PhD in philosophy and works in a philosophy department, but his work focuses more on linguistics than on philosophy. Then (19) is felicitous but (20) is not (Rudolph & Kocurek 2020, p. 7-8):

(19) Al is more a linguist than a philosopher.

(20) #It's more likely that Al is a linguist than a philosopher.

---

[127] See Wellwood (2014) for a defence of a (C)-type account.

5 Conclusion

This chapter provides three contributions to the debate over metalinguistic comparatives. Firstly, it identifies the single-value case (when 'X is F' and 'X is G' are both true and yet 'X is more F than G' is true as well) as an important desideratum for a successful account of MCs. Secondly, it provides a novel interpretation of the sense in which to utter 'X is more F than G' is to say that it's better to say 'X is F' than to say 'X is G'. The account employs the semantic similarity relation derived from the epistemicist account of definiteness developed in Chapter 5. In short, it's better to say that 'X is F' than to say 'X is G' because 'X is F' is more robustly true (or less robustly false). Lastly, this chapter improves the technical work done by Rudolph & Kocurek (2020) by providing truth conditions for metalinguistic comparatives that escape the pitfalls of their account: in particular, issues with the single-value case.

Concluding remarks

My aim in this thesis was to refine the epistemicist treatment of vagueness, as well us to explore semantic plasticity as an independent phenomenon characterising our language. I started by summing up the main points in the debate over epistemicism that ensued after the publication of *Vagueness* by Williamson. The majority of the debate on epistemicism, especially in the decade or so after the publication of *Vagueness*, focused on epistemicist metasemantics; doing justice to this literature required dedicating a separate chapter just to these issues. As I saw it, the biggest threat to epistemicist metasemantics came from the literature on moral vagueness: I aimed to show that epistemicism can take on the challenge.

I think the most significant contribution of the thesis is the refinement of the epistemicist slogan that borderlineness is (associated with) ignorance due to semantic plasticity. I provided a model that allows for basic understanding of the notion of ignorance due to a factor (like semantic plasticity). Furthermore, I analysed and gave formal treatment of the epistemicist notion of definiteness using a particular notion of semantic plasticity, understood as a variation in metasemantic rules in nearby worlds. I tied up loose ends by linking the account of definiteness with a puzzle regarding the semantic plasticity of 'truth' and showing how my account treats such metalinguistic vocabulary. Furthermore, I outlined ways in which epistemicism can be squared with the Omniscient Speaker Puzzle, opting for the error-theoretic

route. Progress is slow and the devil lies in the details. I hope that I have illuminated some of those details and solved some problems along the way.

There are several topics that are naturally connected with the issues I discussed here, that could have also been explored in the thesis. First, my plan was to explore semantic plasticity as a phenomenon independent of the debate on epistemicism. The account of metalinguistic comparatives presented in Chapter 8 is written in that train of thought. However, there are other puzzles that I wish to have covered and solved, e.g. Dorr & Hawthorne's (2014) speech reports puzzle. I will take this to mean that my goal of exploring semantic plasticity as a phenomenon independent of epistemicism was partially achieved; nevertheless, there is still work to be done taking this approach. The consolation is that there are some clear targets for future research.

Second, I think research on semantic plasticity can have some intriguing implications for metaphysics. One example is a puzzle raised by John Hawthorne (2006): semantic plasticity of 'I' and 'person' can result in some intriguing consequences like the vagueness of 'parthood'; another is Ofra Magidor's (2018) argument against the Sider-Lewis argument from vagueness using her account of epistemicist definiteness. As this recent research[128] shows, there is a trend of using the theoretical machinery of semantic plasticity to raise and, more importantly, to answer interesting puzzles in metaphysics. This is another obvious candidate for future research.

Overall, I hope to have shown that the epistemicist project has a promising future. There are multiple paths for the epistemicist to take and explore. Among other topics, the epistemicist faces a choice between the Non-Ambitious and the Ambitious route discussed in Chapter 4, the potential for a switch from safety-based epistemology to normality-based epistemology, an

[128] See also (Dorr, Hawthorne, Yli-Vakkuri 2021).

opportunity to develop a more refined understanding of ignorance relative to a factor. I think these options indicate some exciting opportunities for future developments of epistemicism.

Bibliography

[1] Andjelković, M., & Williamson, T. (2000). Truth, Falsity, and Borderline Cases. *Philosophical Topics*, 28(1), 211–244.

[2] Anscombe, G. E. M., (1957). *Intention*. Basil Blackwell.

[3] Bacon, A., (2018). *Vagueness and Thought*. Oxford University Press.

[4] Ball, B. (2013). Knowledge Is Normal Belief. *Analysis*, 73 (1), 69–76.

[5] Barnes, E., (2009). Indeterminacy, identity and counterparts: Evans reconsidered. *Synthese*, 168, 81–96.

[6] Barnes, E., (2010). Ontic Vagueness: A Guide for the Perplexed. *Noûs*, 44(4), 601–627.

[7] Barnes, E., & Cameron, R. P., (2011). Back to the Open Future. *Philosophical Perspectives*, 25(1), 1–26.

[8] Barnes, E., & Williams, J. R. G. (2011a). A Theory of Metaphysical Indeterminacy. In *Oxford Studies in Metaphysics* Volume 6 (eds. Karen Bennett & Dean Zimmermann) (pp. 103–148). Oxford University Press.

[9] Barnes, E., & Williams, J. R. G. (2011b). Reply to Eklund. In *Oxford Studies in Metaphysics* Volume 6 (eds. Karen Bennett & Dean Zimmermann) (pp. 173–182). Oxford University Press.

[10] Barnes, E., & Williams, J. R. G., (2009). Vague parts and vague identity. *Pacific Philosophical Quarterly*, 90(2), 176–187.

[11] Barnett, D., (2000). Vagueness-Related Attitudes. In *Philosophical Issues* 10: Skepticism (ed. E. Sosa), 302–320. Oxford: Blackwell.

[12] Beddor, B., & Pavese, C. (2020). Modal Virtue Epistemology. *Philosophy and Phenomenological Research*, 101(1), 61–79.

[13] Breckenridge, W., & Magidor, O., (2012). Arbitrary reference. *Philosophical Studies*, 158, 377–400.

[14] Broome, J. (1997). Is Incommensurability Vagueness? In (eds. R. Chang) *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press.

[15] Burgess, J. (2001). Vagueness, Epistemicism and Response-Dependence. *Australasian Journal of Philosophy*, 79(4), 507–524.

[16] Caie, M., (2012). Vagueness and semantic indiscriminability. *Philosophical Studies*, 160(3), 365–377.

[17] Cameron, R. P., (2010). Vagueness and Naturalness. *Erkenntnis*, 72(2), 281–293.

[18] Chang, R. (1998). *Incommensurability, incomparability, and practical reason*. Cambridge, MA: Harvard University Press.

[19] Chang, R. (2002). The possibility of parity. *Ethics*, 112(4), 659–688.

[20] Ciecierski, T., (2022). Demonstrative uses of indexicals. *Warsaw Indexicality Workshop*. University of Warsaw, 6-7 April.

[21] Dorr, C., (2003). Vagueness without Ignorance. *Philosophical Perspectives*, 17, 83–113.

[22] Dorr, C., & Hawthorne, J., (2014). Semantic Plasticity and Speech Reports. *Philosophical Review*, 123(3), 281–338.

[23] Dorr, C., Hawthorne, J., Yli-Vakkuri, J., (2021). *Bounds of Possibility: Puzzles of Modal Variation*. Oxford University Press.

[24] Dougherty, T., (2014). Vague value. *Philosophy and Phenomenological Research*, 89(2), 352–372.

[25] Dunaway, B., & McPherson, T., (2016). Reference Magnetism as a Solution to the Moral Twin Earth Problem. *Ergo*, 3(25), 239–279.

[26] Dutant, J., (2016). How to Be an Infallibilist. *Philosophical Issues*, 26, 148–171.

[27] Eklund, M., (2011). Being Metaphysically Unsettled: Barnes and Williams on Metaphysical Indeterminacy and Vagueness. In *Oxford Studies in Metaphysics* Volume 6 (eds. Karen Bennett & Dean Zimmermann) (pp. 149–172). Oxford University Press.

[28] Evans, G., (1978). Can There Be Vague Objects? *Analysis*, 38(4), 208.

[29] Fine, K., (1975). Vagueness, Truth, and Logic. *Synthese*, 30(3/4), 265–300.

[30] Fritz, P., (2016). Appendix to Juhani Yli-Vakkuri's 'Epistemicism and Modality.' *Canadian Journal of Philosophy*, 46(4–5), 836–838.

[31] Giannakidou, A., & Stavrou, M., (2009). Metalinguistic comparatives and negation in Greek. In Claire Halpert, Jeremy Hartman & David Hill (eds.), *Proceedings of the 2007 workshop on Greek syntax and semantics* (MIT working papers in linguistics 57), 57–74. MIT Press.

[32] Giannakidou, A., & Yoon, S., (2011). The subjective mode of comparison: metalinguistic comparatives in Greek and Korean. *Natural Language & Linguistic Theory* 29, 621–655.

[33] Gómez-Torrente, M., (1997). Two Problems for an Epistemicist View of Vagueness. *Philosophical Issues*, 8(Truth), 237–245.

[34] Gómez-Torrente, M., (2002). Vagueness and Margin for Error Principles. *Philosophy and Phenomenological Research*, 64(1), 107–125.

[35] Goodman, J., (2013). Inexact Knowledge Without Improbable Knowing. *Inquiry*, 56(1), 30–53.

[36] Goodman, J. & Salow, B., (2018). Taking a Chance on KK. *Philosophical Studies*, 175, 183–96.

[37] Graff Fara, D., (2002). An Anti-Epistemicist Consequence of Margin for Error Semantics for Knowledge. *Philosophy and Phenomenological Research*, 64(1), 127–142.

[38] Graff Fara, D., (2003). Gap Principles, Penumbral Consequence, and Infinitely Higher-Order Vagueness. In (eds. J. *Beall) Liars and Heaps: New Essays on Paradox* (pp. 195–222).

[39] Greco, D., (2014). Could KK Be OK? *Journal of Philosophy*, 111(4), 169–197.

[40] Greco, D., (2016). Safety, Explanation, Iteration. *Philosophical Issues*, 26, 187–208.

[41] Hart, W. D., (1992). Hat-Tricks and Heaps. *Philosophical Studies* 33, 1-24.

[42] Hawthorne, J., (2004). *Knowledge and Lotteries*. Oxford University Press.

[43] Hawthorne, J., (2006). Epistemicism and Semantic Plasticity. *Oxford Studies in Metaphysics* 2, pp.289–322.

[44] Horgan, T., & Timmons, M., (1991). New wave moral realism meets Moral Twin Earth. *Journal of Philosophical Research*, 16, 447–465.

[45] Horgan, T., (1997). Deep Ignorance, Brute Supervenience, and the Problem of the Many. *Philosophical Issues,* 8(Truth), 229–236.

[46] Horwich, P., (1997). The Nature of Vagueness. *Philosophy and Phenomenological Research*, 57(4).

[47] Kamp, J. A. W. (1975). Two theories about adjectives. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (pp. 123–155). Cambridge University Press.

[48] Kaplan, D. (1989). Demonstratives. In *Themes from Kaplan*, (eds. Joseph Almog et al). Oxford University Press.

[49] Kaplan, D., 1978. Dthat. In Peter Cole (ed.), *Syntax and Semantics*. Academic Press. pp. 221-243.

[50] Kearns, S., & Magidor, O., (2008). Epistemicism about vagueness and meta-linguistic safety. *Nous-Supplement: Philosophical Perspectives*, 22(1), 277–304.

[51] Kearns, S., & Magidor, O., (2012). Semantic Sovereignty. *Philosophy and Phenomenological Research*, 85(2), 322-350.

[52] Keefe, R., (2000). *Theories of Vagueness*. Cambridge University Press.

[53] Keefe, R., Smith, P., & (eds.). (2010). *Vagueness: A Reader*. MIT Press.

[54] Kocurek, A., Jerzak, E., Rudolph, R., (2020). Against conventional wisdom. *Philosophers' Imprint* 20(22). 1–27.

[55] Kripke, S. (1972). *Naming and Necessity*. Harvard University Press.

[56] Lewis, D., (1973). *Counterfactuals*. Oxford: Basil Blackwell.

[57] Lewis, D., (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61(4), 343–377.

[58] Lewis, D., (1986). *On the Plurality of Worlds*. Blackwell Publishing.

[59] Lewis, D., (1988a). Relevant Implication. *Theoria*, 54, 162–174.

[60] Lewis, D., (1988b). Statements Partly About Observation. *Philosophical Papers* 17, 1–31.

[61] Lewis, D., (1988c). Vague Identity: Evans Misunderstood. *Analysis*, 48, 128–130.

[62] Lewis, D., (1997). Finkish Dispositions. *The Philosophical Quarterly,* 47(187), 143-158.

[63] Lewis, D., (1999). Many, but almost one. In *Papers in metaphysics and epistemology* (pp. 164–182). Cambridge University Press.

[64] Leplin, J. (2007). In Defense of Reliabilism. *Philosophical Studies* 134, 31–42.

[65] Litland, J. E., & Yli-Vakkuri, J., (2016). Vagueness and Modality. *Philosophical Perspectives*, 30(1), 229–269.

[66] Littlejohn, C., & Dutant, J., (2020). Justification, Knowledge, and Normality. *Philosophical Studies*, 177, 1593–1609.

[67] Loets, A., (2022). Choice Points for a Theory of Normality. *Mind* 121(521),159-191.

[68] Lowe, E. J., (1995). The Problem of the Many and the Vagueness of Constitution. *Analysis*, 55(3), 179–182.

[69] Machina, K., & Deutsch, H., (2002). Vagueness, ignorance, and margins for error. *Acta Analytica* 17, 19–45.

[70] Magidor, O. (2018). The Epistemicist Solution to the Sorites Paradox. In S. Oms & E. Zardini (Eds.), *The Sorites Paradox*. Cambridge University Press.

[71] Magidor, O., (2018). Epistemicism, Distribution, and the Argument from Vagueness. *Noûs*, 52(1), 144–170.

[72] Mahtani, A. (2004). The Instability of Vague Terms. *The Philosophical Quarterly*, 54(2017), 570–576.

[73] Manley, D. (n.d.). *Moral realism and semantic plasticity*. Retrieved November 5, 2017, from https://pdfs.semanticscholar.org/1016/d1034b074fe637db653dad3be63a4e9d57bb.pdf

[74] Markosian, N. (1998). Brutal Composition. *Philosophical Studies*, 92(3), 211–249.

[75] McGee, V., & McLaughlin, B., (1995). Distinctions Without a Difference. *The Southern Journal of Philosophy*.

[76] McGee, V., & McLaughlin, B., (1998). Review of Vagueness. *Linguistics and Philosophy*, 21(2), 221–235.

[77] McGee, V., & McLaughlin, B., (2000). The Lessons of the Many. *Philosophical Topics*, 28(1), 129–151.

[78] McGee, V., & McLaughlin, B., (2004). Logical Commitment and Semantic Indeterminacy: A Reply to Williamson. *Linguistics and Philosophy*, 27(1), 123–136.

[79] Merricks, T., (2005). Composition and Vagueness. *Mind*, 114(455), 615–637.

[80] Morzycki, M., (2011). Metalinguistic comparison in an alternative semantics for imprecision. *Natural Language Semantics* 19, 39–86.

[81] Mott, P., (1998). Margins for Error and the Sorites Paradox. *The Philosophical Quarterly*, 48(193), 494-504.

[82] Norcross, A., (2006). The Scalar Approach to Utilitarianism. In Henry West (ed.) *The Blackwell Guide to Mill's Utilitarianism*. Wiley-Blackwell, 217-232.

[83] Peet, A., & Pitcovski, E., (2018). Normal Knowledge: Toward an Explanation-Based Theory of Knowledge. *Journal of Philosophy*, 115(3), 141–157.

[84] Pritchard, D., (2009). Safety-Based Epistemology: Whither Now? *Journal of Philosophical Research* 34, 33-45.

[85] Putnam, H., (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science*, 7, 131-193.

[86] Putnam, H., (1980). Models and Reality. *The Journal of Symbolic Logic*, 45(3), 464–482.

[87] Quine, W. V. O., (1953). Three Grades of Modal Involvement. In (eds. R. F. Gibson) (2004) *Quintessence: Basic Readings from the Philosophy of W. V. O. Quine* (pp. 379–398). Harvard University Press.

[88] Rudolph, R., Kocurek, A., (2020). Comparing conventions. *Proceedings of SALT* 30.

[89] Schiffer, S., (1997). Williamson on Our Ignorance in Borderline Cases. *Philosophy and Phenomenological Research*, 57(4), 937–943.

[90] Schiffer, S., (1999). The Epistemic Theory of Vagueness. *Philosophical Perspectives* 13, 481–503.

[91] Schoenfield, M., (2016). Moral Vagueness Is Ontic Vagueness. *Ethics*, 126, 257–282.

[92] Sennet, A. (2012). Semantic plasticity and epistemicism. *Philosophical Studie*s, 161(2), 273–285.

[93] Shafer-Landau, R., (1995). Vagueness, Borderline Cases, and Moral realism. *American Philosophical Quarterly*, 32(1), 83–96.

[94] Sider, T., (2001). *Fourdimensionalism*. Oxford University Press.

[95] Sider, T., (2011). *Writing the Book of the World*. Oxford University Press.

[96] Smith, M., (2010). What Else Justification Could Be. *Noûs*, 44(1), 10–31.

[97] Smith, M., (2016). *Between Probability and Certainty: What Justifies Belief*. Oxford: Oxford University Press.

[98] Smith, M., (2018a). Coin Trials. *Canadian Journal of Philosophy*, 48(5), 726–741.

[99] Smith, M., (2018b). The Logic of Epistemic Justification. *Synthese*, 195, 3857–3875.

[100] Smith, N. J. J., (2009). *Vagueness and Degrees of Truth*. Oxford University Press.

[101] Sorensen, R., (1988). *Blindspots*. Oxford University Press.

[102] Sorensen, R., (2001). *Vagueness and Contradiction*. Oxford University Press.

[103] Sorensen, R., (2007). Knowledge beyond the Margin for Error. *Mind*, 116(463), 717–722.

[104] Sosa, E., (1999). How must knowledge be modally related to what is known? *Philosophical Topics* 26(1&2), 373-384.

[105] Stalnaker, R., (1978). Assertion. *Syntax and Semantics*, 9, 315-332.

[106] Stalnaker, R., (1999). *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford University Press.

[107] Stalnaker, R., (2005). On Logics of Knowledge and Belief. *Philosophical Studies*, 128, 169–199.

[108] Stalnaker, R., (2015). Luminosity and the KK Thesis. In Sanford C Goldberg. (eds.), *Externalism, Self-Knowledge, and Skepticism: New Essays*. Cambridge: Cambridge University Press.

[109] Sud, R., (2019). Moral Vagueness as Semantic Vagueness. *Ethics* 129, 684–705.

[110] Tarski, A., (1956). The Concept of Truth in Formalized Languages. In *Logic, Semantics, Metamathematics* (pp. 152–278). Oxford University Press.

[111] Thomassen Hjortland, O. (2017). Anti-exceptionalism about logic. *Philosophical Studies*, 174, 631–658.

[112] Tye, M., (1997). On the Epistemic Theory of Vagueness. *Philosophical Issues*, 8(Truth), 247–253.

[113] Unger, P., (1979). There are no ordinary objects. In (eds. D. Graff Fara & T. Williamson) *Vagueness: A Reader*. Ashgate.

[114] Unger, P., (1980). The Problem of the Many. *Midwest Studies In Philosophy*, 5(1), 411–468.

[115] Väyrynen, P. (2019). Normative Commitments and Metanormative Theory. In J. Suikkanen & A. Kauppinen (Eds.), *Methodology and Moral Philosophy*. Routledge

[116] Weatherson, B., (2004). Luminous Margins. *Australasian Journal of Philosophy*, 83, 373–383.

[117] Wellwood, A., (2014). *Measuring predicates*. University of Maryland, College Park PhD dissertation.

[118] Williams, J. R. G., (2006). An Argument for the Many. *Proceedings of the Aristotelian Society*, 106, 411–419.

[119] Williams, J. R. G., (2007). Eligibility and Inscrutability. *Philosophical Review*, 116(3), 361–399.

[120] Williams, J. R. G., (2008a). Ontic Vagueness and Metaphysical Indeterminacy. *Philosophy Compass*, 3(4), 763–788.

[121] Williams, J. R. G., (2008b). Supervaluationism and Logical Revisionism. *The Journal of Philosophy*, 105(4), 192-212.

[122] Williams, J. R. G., (2012). Indeterminacy and Normative Silence. *Analysis* 72(2), 217-225.

[123] Williams, J. R. G., (2014)., Decision-Making Under Indeterminacy. *Philosophers' Imprint*, 14(4), 1-34.

[124] Williams, J. R. G., (2017). Indeterminate Oughts. *Ethics*, 127(3).

[125] Williams, J. R. G., (2018). Normative Reference Magnets. *Philosophical Review*, 127(1), 41–71.

[126] Williams, J. R. G., (2020). *The Metaphysics of Representation*. Oxford University Press.

[127] Williamson, T., (1992). Inexact knowledge. *Mind* 101, 217-242.

[128] Williamson, T., (1994b). *Vagueness*. Routledge.

[129] Williamson, T., (1994a). Definiteness and Knowability. *The Southern Journal of Philosophy*, 33(1 S), 171–192.

[130] Williamson, T., (1994b). *Vagueness*. Routledge.

[131] Williamson, T., (1996). What Makes It a Heap? *Erkenntnis*, 44(3), 327–339.

[132] Williamson, T., (1997a). Imagination, Stipulation and Vagueness. *Philosophical Issues*, 8(Truth), 2015–2228.

[133] Williamson, T., (1997b). Replies to Commentators. *Philosophical Issues*, 8(Truth), 255–265.

[134] Williamson, T., (1997c). Reply to Commentators. *Philosophy and Phenomenological Research*, 57(4).

[135] Williamson, T., (1999). Schiffer on the Epistemic Theory of Vagueness. *Philosophical Perspectives*, 13, 505–517.

[136] Williamson, T., (2000a). *Knowledge and Its Limits*. Oxford University Press.

[137] Williamson, T., (2000b). Margins for Error: A Reply. *The Philosophical Quarterly*, 50(198), 76–81.

[138] Williamson, T., (2002a). Epistemicist Models: Comments on Gómez-Torrente. *Philosophy and Phenomenological Research*, 64(1), 143–150.

[139] Williamson, T., (2002b). Reply to Machina and Deutsch on Vagueness, Ignorance, and Margins for Error. *Acta Analytica*, 17(29), 47–61.

[140] Williamson, T., (2003). Vagueness in Reality. In *The Oxford Handbook of Metaphysics* (eds. M. Loux & D. Zimmermann). Oxford University Press.

[141] Williamson, T., (2004). Reply to McGee and McLaughlin. *Linguistics and Philosophy*, 27, 113–122.

[142] Williamson, T., (2007a). Knowledge Within the Margin for Error. *Mind*, 116(463), 723–726.

[143] Williamson, T., (2007b). *The Philosophy of Philosophy*. Blackwell Publishing.

[144] Williamson, T., (2009). Probability and Danger. *The Amherst Lecture in Philosophy* 4.

[145] Williamson, T., (2016). Reply to Yli-Vakkuri. *Canadian Journal of Philosophy*, 46(4–5), 839–851.

[146] Williamson, T., (2018). Supervaluationism and good reasoning. *Theoria*, 33(3), 521–537.

[147] Williamson, T. (2020). *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford University Press.

[148] Woods, J., (2018). Logical Partisanhood. *Philosophical Studies*, 176, 1203–1224.

[149] Woods, J., (2019). Against Reflective Equilibrium for Logical Theorizing. *Australasian Journal of Logic*, 16(7), 319-341.

[150] Yli-Vakkuri, J., (2013). Propositions and Compositionality. *Philosophical Perspectives*, 27, 526–563.

[151] Yli-Vakkuri, J., (2016). Epistemicism and modality. *Canadian Journal of Philosophy*, 46(4–5), 803–835.