KNOWLEDGE, LIES AND VAGUENESS: A MINIMALIST TREATMENT
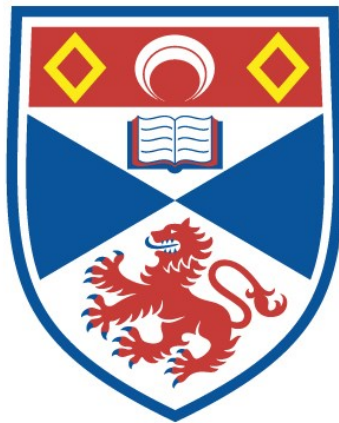
Patrick Greenough

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**

2002

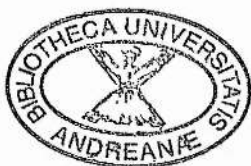# KNOWLEDGE, LIES, AND VAGUENESS

A Minimalist Treatment

by

Patrick Greenough

Thesis submitted for the Degree of Doctor of Philosophy

University of St. Andrews

December 2001

ProQuest Number: 10166931

ProQuest.

ProQuest 10166931

DECLARATIONS

I, Patrick Greenough, hereby certify that this thesis, which is 78 000 words in length, has been written by me, that it is the record of work carried out by me, and that it has not been submitted in any previous application for a higher degree.

Date: 18/6/2002   Signature of candidate:

I was admitted as research student in April 1995 and as a candidate for the degree of Doctor of Philosophy in April 1995; the higher study for which this is a record was carried out in the University of St. Andrews between 1996 and 2001.

Date: 18/6/2002   Signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: Signature of supervisor:

18/6/2002

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date: 18/6/2002   Signature of candidate:

ABSTRACT

Minimalism concerning truth is the view that that all there is to be said concerning truth is exhausted by a set of basic platitudes. In the first part of this thesis, I apply this methodology to the concept of knowledge. In so doing, I develop a model of inexact knowledge grounded in what I call *minimal margin for error principles*. From these basic principles, I derive the controversial result that epistemological internalism and internalism with respect to self-knowledge are untenable doctrines. In the second part of this thesis, I develop a minimal theory of vagueness in which a rigorous but neutral definition of vagueness is shown to be possible. Three dimensions of vagueness are distinguished and a proof is given showing how two of these dimensions are equivalent facets of the same phenomenon. From the axioms of this minimal theory one can also show that there must be higher-order vagueness, contrary to what some have argued. In the final part of this thesis, I return to issues concerning the credentials of truth-minimalism. Is truth-minimalism compatible with the possibility of truth-value gaps? Is it right to say that truth-minimalism is crippled by the liar paradox? With respect to the former question, I develop a novel three-valued logical system which is both proof-theoretically and truth-theoretically well-motivated and compatible with at least one form of minimalism. With respect to the second question, a new solution to the liar paradox is developed based on the claim that while the liar sentence is meaningful, it is improper to even suppose that this sentence has a truth-status. On that basis, one can block the paradox by restricting the *Rule of Assumptions* in Gentzen-style presentations of the sentential sequent calculus. The first lesson of the liar paradox is that not all assumptions are for free. The second lesson of the liar is that, contrary to what has been alleged by many, minimalism concerning truth is far better placed than its rival theories to solve the paradox.

CONTENTS

ACKNOWLEDGEMENTS

I have been told that a student registered for the PhD at the University of St. Andrews can gain their doctorate in one of three (not necessarily exclusive) ways: death, madness, and the production of a thesis of not more than 100 000 words which is expected to constitute a new and significant contribution to learning (consistent with expectations based on what is reasonable within three years of full-time study) and should contain material worthy of publication in some form. At many times, in the writing of this thesis, it has seemed that at least one of the first two routes was by far the more likely road to success. Had it not been for the support and encouragement of many people, madness and death now seem slightly less likely outcomes than they once did.

My wife has been unfailing in her love and encouragement. I do not know what I would do without her. I dedicate this work to her. My mother has been the best mother anybody could wish for and has given unconditional support all the way. My brother Thomas has been an inspiration (not to mention a source of much needed finance). His wife, Michelle, has been mother-earth herself. My nephew Gabriel and his sister Lillian have kept my notice-board and me smiling, and an unborn Greenough will doubtless do the same in a few days time. John Cottam and Jane McQuilin have given untold help. Sven Rosenkranz has marshalled me along and has been the very best of friends. I particularly thank him for reading the whole manuscript and saving me from many errors. Mike Campbell and Paul Brown were a constant source of much needed distraction. Tim Miller showed me how to harvest, which I thank him for greatly. Rosie, Andrew, Mike, and Josie, provided the best holidays one could wish for. James Ladyman, Katherine Hawley, and Fraser MacBride deserve special mention, not least for each keeping the ball rolling. Barney, the cat, purred throughout.

Many thanks also go to Chris Bertram, David Bain, Jessica Brown, Peter Clark, Jimmy Doyle, Keith Graham, Antti Karjalainen, Geoff Leech, Stewart Shapiro, and Mauricio Suárez. The Southgaiters (Lars Binderup, Lars Gunderson, Jesper Kallestrup, Patrice Philie, Duncan Prichard, and Sven Rosenkranz) are incomparable. They are great friends. Without them, this thesis would be barely presentable. I learnt most of my philosophy talking to them. Lots of further acknowledgements are given at the beginning of each chapter. I hope that I have not missed anybody out. Last, but certainly not least, Crispin Wright has been an unbelievable inspiration and has gone well beyond the call of duty in his supervision of what must have seemed, at times, to be a never ending story.

i

INTRODUCTION

The goal of chapters One and Two is to investigate the character of knowledge from a perspective which is as neutral as possible on philosophical matters. In so doing, the aim is to lay the foundation for a *minimal* theory of knowledge. Call this theory *epistemic minimalism*. At the very least, epistemic minimalism promises to ensure that epistemological inquiry can begin at a mutually agreed point. Moreover, should some substantive epistemological theory fail to entail epistemic minimalism then we may dismiss the credentials of this substantive theory from the outset. But epistemic minimalism can also do much more than this. In Chapter Two, it is employed to show that there are good grounds to accept externalism in epistemology, while in Chapter Three, it serves to ground the minimal theory of vagueness.

The overall aim in Chapter Two, then, will be to test whether our minimal axioms are genuinely minimal, discard those that conflict with any further minimal intuitions we may have concerning knowledge, and put the remaining principles to work by seeing if we can derive some fruitful and perhaps controversial theorems from them. As it turns out, though epistemic minimalism is all but based on platitudes, it nonetheless promises to be able to settle certain pressing and long-standing issues in epistemology. In particular, our minimal theory of knowledge proves to be sufficiently rich to refute two key internalist theses, namely: that knowledge iterates (roughly, if one knows and one has properly considered whether one knows then one knows that one knows); and, that the mental realm is transparent to the knowing subject (roughly, if one is in a certain mental state and one has properly considered whether one is in that state then one knows that one is in that state).

Hitherto, there has been a consensus that a constitutive definition of vagueness is too much to ask for. In Chapter Three my aim is to is to challenge that consensus by identifying the constitution of vagueness from a perspective which is as neutral as possible on matters both logical and philosophical. In so doing, I shall lay the foundation for *a minimal theory of vagueness*. To that end, three related dimensions of vagueness are distinguished: vagueness *qua* sorites-susceptibility, vagueness *qua* borderline cases, and vagueness *qua* tolerance. Hitherto, the relationship between these dimensions has remained somewhat unclear. The minimal theory of vagueness is equipped to remove much of that unclarity. One overall merit of this theory is that it promises to ensure that the dialectic of the vagueness debate can at least begin at a mutually agreed point—this theory can at least ensure that we are all talking

about the same phenomenon from the outset. Like the account of knowledge developed in the chapters One and Two, the axioms of this theory are intended to be as uncontroversial as possible while some of its theorems are decidedly controversial. For one thing, it can be shown that, given our minimal axioms, there must be higher-order vagueness.

Is minimalism about truth compatible with truth-value gaps? Is it even possible to have a logic and semantics for truth-value gaps which is both truth-theoretically and proof-theoretically well-motivated? Can truth-value gaps feature in a substantial theory of what it is to be a borderline case? These are the three main questions dealt with in Chapter Four. I develop a novel three-valued logical system which is both proof-theoretically and truth-theoretically well-motivated and compatible with at least one form of minimalism. Is this logic for truth-value gaps of use in a solution to the liar paradox?

Minimalists about truth have had very little to say about the liar paradox. Some have thought that there is a deep reason for this silence—truth-minimalism by its very nature simply lacks any of the right resources to combat the paradox (Simmons 1999). In Chapter Four, it was argued that truth-minimalism is at least compatible with the thesis of truth-value gaps. So can the truth-minimalist employ truth-value gaps in order to solve the liar paradox? According to what Parsons (1984) has dubbed the 'Standard Solution' of the liar paradox, a sentence which says of itself that it is false is a sentence which lacks a truth-value. More sophisticated versions of the Standard Solution take such sentences to be neither *definitely* true nor *definitely* false (McGee 1989, 1991; Soames 1999). The advertised goal of all such proposals is to identify a principled reason to refuse to assert that the liar sentence is (definitely) true/false. It is argued that while the *form* of the Standard Solution is correct, the reasons why a speaker should refuse to assert that the liar sentence is (definitely) true/false have been systematically misidentified hitherto.

An alternative solution (one which retains the shape but not the substance of the Standard Solution) is developed based on the insight that it is improper to even *suppose* the liar sentence to have a truth-status (true or not) on the grounds that supposing a liar sentence to be true/not-true essentially defeats the *telos* of supposition in a readily identifiable way. On that basis, one can block the paradox by restricting the *Rule of Assumptions* in Gentzen-style presentations of the sentential sequent-calculus. The first lesson of the liar paradox turns out to be that not all assumptions are for free.

One key feature of this solution is that it provides a positive argument for a minimalist conception of truth, indeed more specifically a *deflationary* conception of truth in which truth should play no role in our philosophical theorising. There are good reasons for this. Contra

Simmons (1999), it is not that deflationism is worse off than its competitors with respect to solving the liar paradox on the grounds that deflationism cannot employ truth-theoretic resources for substantial philosophical explanations. Rather, it is argued that a deflationary theory of truth is better off than with respect to its competitors. It is the bringing to bear of truth-theoretic resources (such as truth-value gaps) which proves to be problematic and ultimately self-defeating. This is to say that the proposal argued for in this paper is not merely compatible with deflationism, it provides both a positive and a novel reason to accept a deflationary conception of truth. The second lesson of the liar paradox is that deflationism offers the best hope of holding the liar at arm's length.

To return to the beginning, we firstly need to have some idea how epistemic minimalism relates to the more familiar minimalism concerning truth. To that end, in §1.1 a distinction is drawn between three varieties of truth-minimalism: strong, weak, and methodological. In §1.2, I argue that we should at least be methodological minimalists concerning knowledge— that is, we should at least adopt the working hypothesis that all there is to say (of a general nature) concerning knowledge is exhausted by a set of basic platitudes. But what then should we reasonably expect or demand from a minimal theory of knowledge? While epistemic minimalism is not equipped to offer an uncontroversial constitutive definition of knowledge, nor address the threat of scepticism, it can nonetheless set forth some of the necessary conditions on knowledge. But beyond truth and belief what other candidate necessary conditions on knowledge are there?

Williamson (1994, 1997b) has argued that reliability is a necessary condition on knowledge any reasonable conception. But how is such a condition to be expressed? In §1.3, two reliability conditions are considered: the global reliability condition and what I call the relevant connection condition. It is found that the latter condition, while unspecific in nature, offers a reasonably neutral diagnosis as to why a subject may be said to lack knowledge across a range of some but not all Gettier cases. Many authors, including Williamson, assume that the reliability condition can be given a more specific formulation using the resources of modal epistemology. One key virtue of such so-called *local* reliability conditions is that they purport to offer a diagnosis of why a subject lacks knowledge in *all* types of Gettier case.

Four candidate local reliability conditions are considered: sensitivity, stability, safety and robustness. The sensitivity conditions and stability conditions are the familiar tracking conditions given by Nozick (1981). The sensitivity condition is rejected as a minimal

condition as it is too concessive to the sceptic, while the stability condition is rejected as it conflicts with (minimal) intuitions concerning the possibility of lucky knowledge— knowledge that we might easily not have had. The safety condition has been defended by Williamson (1994, 2000), Sosa (1996, 1999, 2000), and Sainsbury (1997). Very roughly, this condition says that knowledge that p requires that one does not form a false belief that p in close cases. On the face of it, this seems like the sort of modal condition that can happily feature as an axiom in our minimal theory. Even if that is so, it is argued that safety alone is not enough. There are well-known Gettier cases where a subject purportedly lacks knowledge but where the safety condition, and indeed the relevant connection condition, are both satisfied. This suggests that we need a further condition, namely *robustness*. This condition says that knowledge that p requires that one does not form the false belief that not-p in close cases. Provisionally at least, there are grounds to think that our minimal theory is committed to the view that knowledge requires both safety and robustness. Lastly, in §1.5, I assess whether there are any good arguments for the belief condition on knowledge. It is found that there are only very weak grounds to accept the belief condition as a minimal condition on knowledge, but that there are at least grounds to accept the principle which says that knowledge that p (gained via method M) entails that one does not believe (via M) that not-p.

Williamson (1992a, 1994, 2000) has employed the safety requirement on knowledge to serve as the foundation for a general model of what he terms *inexact* knowledge. Since this model is derivable from our minimal axioms, plus other apparently acceptable principles, then it ought to feature in our minimal theory of knowledge. In §2.1-§2.4, Williamson's model of inexact knowledge is evaluated in detail and its minimalist credentials are in fact found wanting. One main feature of the discussion will be that it is simply mistaken to think that knowledge requires a local reliability condition.

In §2.1, we look at how Williamson uses his model of inexact knowledge, and particular his *margin for error principles*, to undermine the principles of iterativity and transparency mentioned above. These margin for error principles tell us that knowledge that p requires that it is not an easy possibility that *not-p*. In §2.2, I show how Williamson derives these principles from the thesis that knowledge requires safety, plus a key doxastic principle (roughly, the principle that there is limited discrimination in the belief-forming process). In §2.3, the pedigree of this doxastic principle is questioned on two counts: this principle does not comport well with Williamson's *knowledge-first* methodology, and it is not obvious that this principle is necessary or knowable *a priori*. This latter worry impinges on whether one can exploit Williamson's margin for error principles in support of a realist conception of

truth—this is an issue which I return to in §3.8. In §2.4, I assess whether it is right to say that knowledge requires a (local) reliability condition—either in the guise of Williamson's safety requirement or in the guise of something like Nozick's more familiar tracking conditions. It is found that the phenomenon of lucky knowledge—knowledge we might easily not have had—requires us to reject *all* forms of the local reliability requirement. Hetherington (1998) has further argued that a subject has lucky knowledge in *all* Gettier cases where this subject has a justified true belief. It is argued that while Hetherington is right to posit lucky knowledge in the well-known Henry-barn-facade case and the Jill-assassinated-dictator case, he is wrong to posit lucky knowledge in the Smith-Nogot-Ford existential generalisation case and similar cases. In §2.5, a case is made for the instatement of a causal version of the relevant connection condition on knowledge encountered in Chapter One in place of any local reliability condition. This condition allows us to distinguish Gettier cases where we have lucky knowledge from Gettier cases where we lack knowledge. Some of the problems and prospects for a causal theory of knowledge are evaluated and a case is made for rehabilitating what has long been thought to be a defunct view.

Hetherington has also argued that cases of lucky knowledge show that we should reject the project of modal epistemology altogether. Very roughly, Hetherington takes the lesson of lucky knowledge to be that whether one knows does not depend in any way on what conditions obtain in nearby possible worlds. While Hetherington is right to reject local reliability as a condition on knowledge, he is wrong to reject the ambitions of modal epistemology. Modal epistemology and lucky knowledge are compatible if one suitably weakens the modal conditions for knowledge. This insight informs the minimal theory of knowledge developed in §2.6. This theory utilises what I term *minimal margin for error principles*, principles which tell us that knowledge that p requires that it is not an easy possibility that one knows that *not-p*. Given these basic modal principles, we can nonetheless still show that the internalist principles of iterativity and transparency given above both fail. So, from axioms which are at least *prima facie* acceptable to everyone, it follows that we are all externalists now.

But what can we reasonably expect or demand from a minimal theory of vagueness? Can this theory solve the sorites paradox? Can it isolate the source of linguistic vagueness? Can this theory successfully rehabilitate the so-called *characteristic sentence approach* to defining vagueness? These are the sorts of questions addressed in §3.1. In §3.2, it is found that vagueness defined as sorites-susceptibility offers the least controversial characterisation of vagueness. However this characterisation proves to be too insubstantial for the promises of

the minimal theory to be properly satisfied. On what is perhaps the most prevalent conception, vagueness is the phenomenon of borderline cases. From §3.3 through to §3.5, I assess whether it is plausible to give an uncontroversial definition by reference to such a phenomenon. A number of non-epistemic and epistemic accounts of what it is to be a borderline case are scrutinised. For the purpose of finding a neutral definition of vagueness, none of these proves entirely satisfactory (the particular bug-bear proves to be the possibility of terms which we can stipulate to give rise to borderline cases but which draw sharp and clearly identifiable divisions across their associated dimension of comparison). *Prima facie*, it is far more plausible to minimally define vagueness by reference to an epistemic notion of *tolerance*. Such a notion is intended to capture the thesis that vague terms draw no clear or known boundary across their range of signification and contrasts sharply with the (semantic) notion of tolerance given by Wright (1975, 1976). This suggestion, which relates to insights encountered in Chapter Two, is pursued in §3.6. In §3.7, it is shown that vagueness *qua* borderline cases (when properly construed so as to exclude terms we are stipulated to give rise to borderline cases) and vagueness *qua* epistemic tolerance are in fact conceptually equivalent, contrary to what might be expected.

A puzzle left over from Chapter Two, §2.3, is whether the minimal theory of knowledge entails a realist conception of truth. A similar worry applies to the minimal theory of vagueness, for it looks as if this theory entails that the truth-values of borderline statements are unknowable. However, in §3.8, it is found that while the minimal theory of vagueness, and the minimal margin for error principle **MME**, do indeed entail the existence of undetectable truth, this fact lends no support whatsoever to realism. The principle **MME** is available to everyone. Lastly, in §3.9, I re-configure the anti-internalism argument employed at the end of Chapter Two to show, that there must be higher-order vagueness, contrary to what some have argued.

In §4.1, I set forth a range of arguments which allegedly reveal the truth-theoretic and proof-theoretic weaknesses of any 'gappy' logic. In §4.2, it is argued that truth-minimalism and truths-value gaps are perfectly compatible. One key thought here is that the so-called transparency property of truth does not hold in full generality (as Wright and others have argued). If the arguments in this section are correct, it looks like we should favour a deflationary, non-explanatory form of minimalism. In §4.3, I develop a logic and semantics for truth-value gaps which is both truth-theoretically and proof-theoretically well-motivated (despite the many doubts raised by Williamson, Machina, Horwich, and others). One key move in this section is to express the T-schema using a non-contraposible extensional

conditional allowing the deduction theorem to be retained in full generality. Though most of the given proof-theory is available to the supervaluationist, the rule of disjunction-elimination is taken to be valid and so the semantics for disjunction is truth-functional. (With respect to vagueness, the penumbral connections are thus not-validated.) A novel three-valued truth-functional logic results, one which is arguably more satisfactory than any competitor three-valued logic.

In §4.4, I offer some reasons why the penumbral connections are not validated. It is argued that our intuitions as to whether or not the penumbral connections should be sanctioned by any respectable theory of vagueness are in any case hampered by an important, though entirely neglected, challenge given by Wright (1995). The nub of this challenge is that indeterminacy in truth-value (*qua* borderline case vagueness) ought to be a status compatible with the poles of truth and falsity—a challenge that, if correct, would rule out any three-valued, many-valued, or supervaluational conception of vagueness from the start. In reply to this challenge, it is argued that one can both respect Wright's challenge *and* find a place for truth-value gaps in a theory of vagueness, should one wish to do so.

CHAPTER ONE
EPISTEMIC MINIMALISM[1]

The goal of this chapter and the next is to investigate the character of knowledge from a perspective which is as neutral as possible on philosophical matters. In so doing, the aim is to lay the foundation for a *minimal* theory of knowledge. Call this theory *epistemic minimalism*. At the very least, epistemic minimalism promises to ensure that epistemological inquiry can begin at a mutually agreed point. Moreover, should some substantive epistemological theory fail to entail epistemic minimalism then we may dismiss the credentials of this substantive theory from the outset. But epistemic minimalism can also do much more than this. In Chapter Two, it is employed to show that there are good grounds to reject both internalism with respect to self-knowledge and internalism with respect to knowledge in general. While in Chapter Three, it serves to ground the minimal theory of vagueness.

As a preliminary to such investigations, we firstly need to have some idea how epistemic minimalism relates to the more familiar minimalism concerning truth. To that end, in §1.1 a distinction is drawn between three varieties of truth-minimalism: strong, weak, and methodological. In §1.2, I argue that we should at least be methodological minimalists concerning knowledge—that is, we should at least adopt the working hypothesis that all there is to say (of a general nature) concerning knowledge is exhausted by a set of basic platitudes. But what then should we reasonably expect or demand from a minimal theory of knowledge?

---

[1] This chapter and the next are both based on talks given at the St. Andrews Philosophy Department Study Retreat held at Raasay House, Raasay, Skye, Scotland, May 25th-29th, 2001, and at the Research Seminar, Department of Philosophy, University of Bristol, 28th November, 2001. For very useful feedback, I am particularly indebted to Roy Cook, Katherine Hawley, Agustin Rayo, Sven Rosenkranz, Stewart Shapiro, and John Skorupski on the former occasion, and to Chris Bertram, Andrew Harrison, James Ladyman, and John Mayberry on the latter. Thanks also to Duncan Pritchard and Sven Rosenkranz for detailed written feedback on an earlier draft of Chapters One and Two.

While epistemic minimalism is not equipped to offer an uncontroversial constitutive definition of knowledge, nor address the threat of scepticism, it can nonetheless set forth some of the necessary conditions on knowledge. But beyond truth and belief what other candidate necessary conditions on knowledge are there?

Williamson (1994, 1997b) has argued that reliability is a necessary condition on knowledge any reasonable conception. But how is such a condition to be expressed? In §1.3, two reliability conditions are considered: the global reliability condition and what I call the relevant connection condition. It is found that the latter condition, while unspecific in nature, offers a reasonably neutral diagnosis as to why a subject may be said to lack knowledge across a range of some but not all Gettier cases. Many authors, including Williamson, assume that the reliability condition can be given a more specific formulation using the resources of modal epistemology. One key virtue of such so-called *local* reliability conditions is that they purport to offer a diagnosis of why a subject lacks knowledge in *all* types of Gettier case.

Four candidate local reliability conditions are considered: sensitivity, stability, safety and robustness. The sensitivity conditions and stability conditions are the familiar tracking conditions given by Nozick (1981). The sensitivity condition is rejected as a minimal condition as it is too concessive to the sceptic, while the stability condition is rejected as it conflicts with (minimal) intuitions concerning the possibility of lucky knowledge—knowledge that we might easily not have had. The safety condition has been defended by Williamson (1994, 2000), Sosa (1996, 1999, 2000), and Sainsbury (1997). Very roughly, this condition says that knowledge that p requires that one does not form a false belief that p in close cases. On the face of it, this seems like the sort of modal condition that can happily feature as an axiom in our minimal theory. Even if that is so, it is argued that safety alone is not enough. There are well-known Gettier cases where a subject purportedly lacks knowledge but where the safety condition, and indeed the relevant connection condition, are both satisfied. This suggests that we need a further condition, namely *robustness*. This condition says that knowledge that p requires that one does not form the false belief that not-p in close cases. Provisionally at least, there are grounds to think that our minimal theory is committed to the view that knowledge requires both safety and robustness.

Lastly, in §1.5, I assess whether there are any good arguments for the belief condition on knowledge. It is found that there are only very weak grounds to accept the belief condition as a minimal condition on knowledge, but that there are at least grounds to accept the principle which says that knowledge that p (gained via method M) entails that one does not believe (via M) that not-p.

Minimalists concerning truth generally agree that there is not much to be said about truth beyond the fact that an assertion that p does not differ in any significant way from an assertion that p is true. Moreover, on that basis, they also agree that the theory of truth is not particularly deep or interesting. Beyond that, minimalism fragments into a motley of related but largely distinct theses. Under the rubric of minimalism one typically finds, for instance, the following slogans: truth is not a property; the truth-predicate is redundant; truth has no hidden essence; the truth-predicate merely denotes a logical property; the truth-predicate has an expressive but no theoretical role; the theory of truth is explanatorily lightweight; the truth-predicate merely has a thin conceptual role; that the theory of truth is exhausted by the T-schema; and so on.[2] Rather than attempt to disentangle and adjudicate between these various strands of minimalism here, it merely pays at this point to distinguish *three* basic forms of minimalism about truth.

Call *strong* minimalism concerning truth the doctrine that all there is to be said concerning the concept of truth is completely given to us by a set of *a priori*, conceptually basic and platitudinous principles. In the hands of Horwich (1990, 1998a), this is simply cashed out as the claim that the following schema ('the equivalence schema') implicitly defines the concept of truth:

(**ES**)   The proposition that p is true if and only if p

For Horwich, one possesses the concept of truth just in case one can recognise the validity of all instances of **ES**.[3] Of course, the strong truth-minimalist is equally free to let the truth-bearers be utterances (that say that something is the case), in which case truth would be defined by all instances of the following disquotational schema:

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

Or, equally, one could let the primary truth-bearers be declarative sentences, and maintain that to grasp the concept of truth is just to grasp all instances of Tarski's T-schema which may be given as follows:

---

[2] See O'Leary Hawthorne and Oppy (1997) for a useful survey of the different facets of truth-minimalism

(TS)   *S* is true if and only if *p*

(where '*S*' is the quotation name for the sentence which says that *p*).[4] In the hands of Wright (1992a, 1999), strong minimalism becomes the doctrine that the essential and basic features of truth are exhausted not simply by **ES** (or by **DS/TS**) but by a more comprehensive set of *a priori* platitudes, of which **ES** (or **DS/TS**) are central. To understand the concept of truth is just to grasp the veracity of all these platitudes.[5]

Horwich and Wright also differ in a further key respect. Wright maintains that Horwich's minimalism, and related forms, are unstable since truth records a norm distinct from that of warranted assertibility.[6] So, while Horwich takes truth to be a property which cannot be employed in significant philosophical explanation, Wright makes room for a explanatory role for truth despite the fact that all there is to say about truth can be specified from within a theoretically lightweight perspective (1992a, p. 38). Given this, it thus pays to distinguish a *deflationary* version of strong minimalism, whereby truth has no explanatory role, and an *inflationary* version of strong minimalism, whereby it is allowed that truth may have a part to play in at least some philosophical explanations.[7]

---

[3] Horwich (1998a, pp. 41-2) recognises that the equivalence schema must be appropriately restricted in order to accommodate the semantic paradoxes. This is an issue to which we return to in Chapter Five.

[4] For present purposes, I shall remain neutral as to whether the primary truth-bearers are utterances which represent the world as thus and so, declarative sentences which says that something is the case, or propositions. See Field's (1992) review of Horwich's *Truth* for some relevant discussion.

[5] Wright offers the following list of truth-platitudes: truth does not admit of a more or less (absoluteness); truth is distinct from justification (contrast); any attitude towards a proposition is an attitude towards the truth of that proposition (transparency); truth is a property that is never lost (timelessness); a true proposition corresponds to reality (correspondence); that a truth-apt content remains truth-apt both within the scope of negation (and other sentence-functors) and within the scope of a propositional attitude (embedding); and that some truths may never be known (opacity) (1999, pp. 226-7; see also 1992a, p.34). It is also part of Wright's view to accommodate a moderate pluralism concerning truth in that the place-holders in the truth-platitudes for the target concept may be realised by more than one concept (1992a, p.75; 1999, p. 228).

[6] Wright 1992a, Ch.1. Very roughly, Wright's so-called 'inflationary argument' runs as follows: even though the deflationist thinks that the truth-predicate has no *theoretical* role to play it is nonetheless conceded that it at least has an *expressive role* such that a speaker can employ this predicate to record the fact that a certain statement is warrantedly assertible. (As in the case of 'Everything Osama Bin Laden says is true'.) But while the predicates 'is warrantedly assertible' and 'is true' do not differ in positive normative force, they do differ in content under the scope of negation. Thus, one can allow that neither *p* nor its negation is warrantedly assertible, but one cannot allow that neither *p* nor its negation is true since, for, given the T-schema and classical contraposition, this issues in a contradiction (see Chapter Four for more details on the logic of truth-value gaps). Consequently, truth must record a distinct norm over and above that of warranted assertibility. There is thus more to the predication of truth than the expressive role posited by the deflationist.

[7] It's the ambition of Wright to employ his version of minimalism to undermine certain philosophical theories from the outset. Mackie's error-theory is one such theory (Mackie 1977). (Unlike O'Leary-Hawthorne and Oppy (1997), I shall not treat 'deflationism' and 'minimalism' as interchangeable: deflationism entails minimalism but not vice versa.)

Call *methodological* minimalism concerning truth the doctrine that one should at least begin by assuming strong minimalism as a working hypothesis. Such a position leaves it open whether the constitution of truth is exhausted by a set of basic platitudes. After due philosophical inquiry, it may well turn out that our theory of truth ought to contain some substantial and controversial principles. This may occur when it is recognised that a strong minimal theory of truth proves to be too insubstantial to give a proper theory of, for example, meaning, vagueness, and the liar paradox. Like strong truth-minimalism, methodological truth-minimalism has a deflationary and an inflationary variety. In the former case, one simply assumes that one's theory of truth is exhausted by the equivalence schema **ES** (or by **DS/TS**). On this view, we presume that truth cannot feature in substantial philosophical explanations. In the latter case, one provisionally accepts that something like Wright's list of truth-platitudes does indeed exhaust all there is to say about truth while nonetheless allowing that truth may have a theoretical role to play. The inflationary variant of methodological truth-minimalism says that we should at least begin with Wright's truth-platitudes in order to see how far one is then able to meet the conceptual or explanatory demands placed upon one's theory of truth.[8] Field (1994a, 2000) has effectively defended a deflationary version of methodological minimalism.

It may well turn out that our philosophical inquiries eventually lead us to adopt what may be termed *weak* truth-minimalism. Call weak minimalism about truth the view that the nature of truth is not exhausted by the truth-platitudes, but that these principles nonetheless remain central to our theory of truth such that any respectable theory of truth must at least *entail* these platitudes. As might be expected, the deflationary variant of weak truth-minimalism is the view that there is only one truth platitude (**ES** or **DS** or **TS**) and that this platitude itself will have no role to play in philosophical explanation. The inflationary variant of this view, in contrast, allows that the list of truth-platitudes given by Wright, while not exhausting all there is to say about truth, are central and can nonetheless feature in some substantial philosophical explanations. It's fair to say that Tarski (1944) offered something like the deflationary version of weak minimalism since he held that every substantial theory of truth must be materially adequate in that it should entail all instances of his T-schema **TS**.[9]

---

[8] At the very least a theory of truth must give us an account of the function of the truth-predicate together with a theory of the meaning of the word 'true' (see Horwich 1998a, pp. 36-7).

[9] For relevant discussion see Kirkham (1992, Ch. 6).

We shall not engage with the topic of truth-minimalism again until the last chapter.[10] For present purposes, I merely wish to ask: can one and should one be a minimalist about knowledge?

1.2 Epistemic Minimalism

Given the discussion in the previous section, it should come as no surprise that we can distinguish three basic forms of what may be termed *epistemic* minimalism. Call *strong* epistemic minimalism the view that one can *exhaust* all there is to say about knowledge by specifying a set of *a priori* and basic platitudes; call *methodological* epistemic minimalism the view that one should at least adopt the working hypothesis that strong epistemic minimalism is correct; and call *weak* epistemic minimalism the view that at least part of the nature of knowledge cannot be specified from within a minimal theory. Which species of epistemic minimalism should we prefer?

It might seem obvious that weak epistemic minimalism is true and seem obvious that strong generic minimalism is false. Surely everybody agrees that knowledge entails truth. Thus, there is at least one platitude concerning knowledge—the truth-platitude. But is there more than one—and in particular, is there a set of platitudes which exhausts the content of 'S knows that *p*'? It certainly has to be conceded that there is very little agreement as to what further necessary conditions one should place on knowledge, let alone any agreement as to what are the sufficient conditions for when a subject counts as knowing. For this reason, it might seem that the prospects for a strong minimalist approach to knowledge are rather bleak—much bleaker than any hope for a strong minimal account of truth. Thus, there are

---

[10] Are there are any *prima facie* reasons to adopt one form of truth-minimalism rather than another? Though there is not space to deal with this question in any detail, Wright's inflationary argument gives one initial reason to adopt some inflationary form of minimalism. Once one accepts the platitude that truth and warranted assertibility are distinct then there is reason to believe that there is indeed more to truth than the expressive role posited by the deflationist. But which form of inflationary minimalism should be preferred—weak, strong, or methodological? There are a number of reasons why one might favour weak over strong (inflationary) truth-minimalism. In the first place, Wright not only concedes that some of his platitudes might be questioned, but also that his list may well be incomplete (1992a, pp. 72-3; 1999, p. 227). While its fairly easy to augment Wright's list, we are nonetheless given no guarantee that whatever plausible list we produce *absolutely* exhausts all there is to say about truth. For all we can presently say, the extant platitudes at best provide a partial characterisation of truth which may require supplementing by further controversial principles. Moreover, even if such a list were exhaustive, we would only have succeeded to have fixed the content of truth insofar as we have laid down further sets of platitudes which fix all the other concepts mentioned in each of the truth-platitudes, i.e. the concepts of assertion, denial, belief, judgement, for instance. Given this predicament, it seems best to adopt the methodological variant of inflationary truth-minimalism. The spirit of our investigations will thus be exploratory. It is not until the final chapter that we will have reason to dispense with methodological truth-minimalism and adopt a view which is more deflationary in character.

grounds to say that that weak epistemic minimalism is correct from the outset. Such pessimism is unjustified. For one thing, it may well turn out that a strong minimal conception of knowledge will exhaust all there is to say of a *general* nature concerning the concept of knowledge, but without thereby giving a complete characterisation of this concept.[11] At the start of inquiries we cannot rule such a possibility out. Consequently, we should adopt methodological epistemic minimalism and hence endorse strong epistemic minimalism as a working hypothesis. Such issues in any case raise the question of just what we should expect or demand from a minimal theory of knowledge?

Should we demand that epistemic minimalism (weak or strong) be deflationary character? We should certainly not rule out that a minimal theory of knowledge has no work to do in philosophical explanation. Arguably, a deflationary variant of epistemic minimalism (weak or strong) is intelligible but wrong-headed. Consider the truth-condition on knowledge. Surely this platitude can feature in substantial epistemic explanations for it can be employed in explanations as to why a subject lacks knowledge: If James knows that Jimmy is alive then Jimmy is alive, but Jimmy is not alive, therefore James does not know that Jimmy is alive. Michael Williams seems to defend one form of deflationary epistemic minimalism when he says:

> A deflationary account of 'know' may show how the word is embedded and in a teachable and useful linguistic practice, without supposing that 'being known to be true' denotes a property that groups propositions into a theoretically significant kind (1996, p. 113).

Does this mean that there is nothing to say about the concept of knowledge? As I read him, Williams seems to think that there is something to say concerning the concept of knowledge. A theory of the concept of knowledge will typically contain 'highly formal remarks' (what I am here calling platitudes) about this concept, but without thereby specifying substantive and specific constraints on the conditions for when a sentence counts as known.[12] We should thus not read Williams as excluding the possibility that there are platitudes which hold true of the

---

[11] This relates to the so-called *generality problem* faced by reliabilist conceptions of knowledge (see Conee and Feldman 1998). Suppose it is a general and uncontroversial claim that knowledge requires reliability. This general claim falls short of telling us just how much reliability knowledge requires in particular cases. It seems unlikely that any general formula could be given since the degree of required reliability will vary from context to context. Thus, in grasping all the general platitudes there may be governing knowledge one does not necessarily thereby demonstrate a mastery of the expression 'S knows that *p*' (*cf.* Williamson 2000, p. 100).

[12] One reason Williams advances in favour of such a 'deflationary' view of knowledge is that the expression 'S knows that *p*' is context-sensitive. One could even advance the view that it is a platitude that this expression is context-sensitive. In that sense, we are all epistemic contextualists now. However, Williams also wishes to reject what he calls 'epistemological realism'—the view that there are no 'invariant epistemological constraints underlying the shifting standards of everyday justification (1996, p.113). This Wittgensteinian theme seems to go too far. There is one invariant rule in every context, namely: 'Don't doubt the hinge proposition'.

concept of knowledge from which we can derive surprising and perhaps controversial theorems. This point is particularly relevant in Chapter Two when it is shown that surprising results can be obtained from the axioms of epistemic minimalism. But then the following question arises: is it not illicit to allow that our minimal theory of knowledge should permit the derivation of controversial theorems—if the theorems are controversial (and our background logic is not open to question) then surely our axioms should likewise be controversial?

This worry is accommodated by the methodological character of our investigations. Our aim is to offer a set of axioms which on the face of it seem to chime with our ordinary thinking about knowledge.[13] The *prima facie* case is that these axioms ought to be accepted by everyone. Should we then be able to derive theorems which do not chime with everybody's conception of knowledge then, *prima facie*, so much the worse for their conception of knowledge. The burden of proof then falls on the theorist who disagrees with one or more of the theorems to find a relevant failing in what initially seemed to be platitudinous principles. Should no relevant failing be forthcoming, then we are entitled to reject their conception of knowledge outright. With these particular worries out of the way, should we demand that our minimal theory of knowledge address the traditional goals of epistemology?

There are two main traditional goals of epistemology: to furnish some workable definition of knowledge and to undermine the threat of scepticism. These aims are not unrelated. Set the conditions for knowledge too high and scepticism inevitably results; set the conditions too low and one will fail to fix the content of the concept in any relevant way. Even if we do not expect of our minimal theory of knowledge that it be able to combat scepticism in any substantial way, it's tempting to demand that such a theory ought to nonetheless yield a generally acceptable definition of knowledge.[14] In one sense that's an easy demand to meet.

---

[13] In developing his truth-minimalism, Wright thinks that the truth-platitudes should 'chime with our ordinary thinking about truth' (1992a, p.72). This does not entail that a minimal theory of truth should simply be a compendium of all the principles concerning truth which philosophers just happened to have agreed on. Perhaps they have disputed principles which they should not have done and not disputed principles which they should have done. Hence there is scope in a minimal theory of truth to reformulate, assess, and dispute just which principles really are the axioms of the theory. The same holds true for the minimal theory of knowledge sketched here. Moreover, there are grounds to think that one should at least be a methodological minimalist concerning every philosophical concept; but that is a claim I cannot substantiate here.

[14] Williams thinks that one can advance beyond giving a minimal theory of *the concept* of knowledge to giving what he calls a *theory of knowledge itself* without invoking any inflationary commitments (such as the commitment to there being a foundation to knowledge). Such a minimal theory of knowledge may well have the resources to address scepticism and other substantive epistemological questions. He takes Austin and Wittgenstein to have offered minimal theories of knowledge which go beyond the mere attempt to state general

One might simply offer the following characterisation: knowledge is a state of believing a true proposition in a good way (see Zagzebski 1999). But then what is it to believe in a good way? The demand then becomes to give some more specific definition of knowledge which gives a better explanation of what counts as good believing. That's a demand that is not easily met by any theory, minimal or otherwise. The familiar stumbling block is the possibility of Gettier-style counterexamples to any tripartite analysis of knowledge in terms of truth, belief, and justification (or justification-surrogates).

But what did Gettier show exactly? We can all agree that Gettier demonstrated that a speaker S may have a justified true belief which is not sufficient for knowledge, at least where S's justification transmits across entailments that S is justified in believing, and S's justification is fallible, i.e. not necessarily truth-entailing. But there is little agreement on just *why* a speaker lacks knowledge in the standard Gettier cases, and indeed there is little agreement as what are the necessary and sufficient conditions for when a speaker's epistemic predicament counts as being a standard Gettier case.[15] One resolution of the Gettier problem (in all its many guises) is to demand that justification be simply defined as whatever it is that makes the difference between true belief and knowledge.[16] On that basis, can one retain the JTB analysis of knowledge and allow each substantive theory of knowledge to specify in detail what is meant by justification based on considerations which are local to that theory?

Since justification has been characterised by reference to knowledge then this definition is clearly circular. But this ought to present no particular worry for the minimalist. Knowledge may in the end admit of no genuinely reductive non-circular analysis, though the definition as given does not entail that such a reductive analysis will never be forthcoming.[17] A greater worry is that when justification is defined as whatever it is that makes the difference between

---

features of the concept of knowledge. I shall merely be concerned with the concept of knowledge and shall bracket issues concerning scepticism.

[15] This is something which I shall return to again below. The standard Gettier cases I take to be cases (like Gettier's own Smith–Jones case and Jones–Ford case) which require both justification to transmit across justified entailments and justification to be fallible. As is well known, the transmission principle is not essential to all Gettier type cases. Chisholm (1966, p. 23, fn.22) seems to have been the first to recognise this, as in his case of having a justified true belief that *there is a sheep in the field* based on the false belief that *I see a sheep in the field*. Here, there is no logical entailment from 'I see a sheep in the field' to 'There is a sheep in the field'; rather, this inference relies on a method of belief formation which allows us to infer from sheep-appearances to the presence of sheep.

[16] This is roughly the tactic of Plantinga (1993, 1995) and Merricks (1995), though these authors prefer to use the term 'warrant' rather than 'justification'—the latter term being reserved for a specifically internalist notion of warrant.

[17] Williamson (2000) has recently (and persuasively) argued that knowledge is not amenable to a reductive or genuinely conjunctive analysis. Note that even if justification could be characterised independently of knowledge, a JTB-account of knowledge would nonetheless remain circular if, as Williamson thinks, it is constitutive of belief that the *telos* of belief is knowledge, since belief is here defined by reference to knowledge (2000, pp. 266-9).

true belief and knowledge, there is no scope for a fallibilist conception of justification.[18] Since it is certainly not a platitude that justification is infallible (i.e. truth-entailing) then it follows that we simply cannot accept a minimal JTB analysis of knowledge where justification is taken to be what makes the difference between true belief and knowledge. Interestingly enough, this predicament in any case chimes with recent developments in epistemology.

One response to the difficulty of furnishing a constitutive and substantial definition of knowledge is to dilute the traditional definitional goal of epistemology (see Morton 2000, section II). How should one do this? Plantinga (1993, 1995), for one, has doubted whether knowledge is amenable to a constitutive characterisation in terms of necessary and sufficient conditions. He thinks that knowledge may at best may receive a more rough and ready analysis in terms of defeasible or criterial conditions. On the other hand, Williamson (2000) has recently urged that the search for a reductive or genuinely interesting conjunctive analysis of knowledge is deeply misplaced. For Williamson, knowledge is both conceptually and explanatorily prior to such concepts as belief and justification. Nonetheless, many authors, including Williamson, Sainsbury (1997), and Sosa (2000), are still willing to set forth what they take to be some of the necessary conditions of knowledge. On this approach, there are general principles which allow us to predict when a subject lacks knowledge, but it is left open whether there are any general principles which allows us to predict when a subject has knowledge.[19] In the next section, we will see if one can incorporate their insights into the minimal theory of knowledge.

---

[18] See Merricks (1995). We must take care to be clear what is meant by infallibilism/fallibilism (something Merricks does not always do). Take infallibilism to be the view that warrant or justification entails truth, such that fallibilism is the view that one can have a justified but false belief. Relevant Alternative theorists (such as Dretske 1970, Goldman 1976, Cohen 1988, and Lewis 1996) define infallibilism to be the view that to be warranted in believing *p* one must rule out *all* the alternatives incompatible with *p*, and fallibilism as the view that to be warranted in believing *p* one merely needs to rule out the *relevant* alternatives to *p*. But note that if 'ruling out' means 'knowing' then one can be an fallibilist in the relevant-alternatives sense and yet an infallibilist in the sense that warrant or evidence is truth-entailing. How so? Actuality is a relevant alternative (if anything is). Hence, if having a warrant for *p* entails ruling out all relevant not-*p* possibilities then to have a warrant for *p* entails that one rules out, that is *knows*, that not-*p* does not obtain in the actual world. It follows that *p* does obtain. Hence, warrant entails truth. However, if warrant entails a weaker non-factive form of *ruling out* not-*p* possibilities then warrant will not entail truth. Arguably, Austin (1946) was a fallibilist in both senses of the term.

[19] On this approach, there is no attempt to solve the Gettier problem by augmenting or refining the tripartite analysis of knowledge. Nonetheless, Gettier cases remain of interest, for these approaches are required to explain why we intuitively lack knowledge across a range of Gettier cases. In giving such an explanation, there is no demand that any candidate necessary condition on knowledge should also form part of some set of jointly sufficient conditions.

If anything is a platitude governing knowledge it is the truth-condition, which we give as follows:

(TC)   Necessarily, if a subject S knows that *p* then *p* is true

Principle **TC** just says that knowledge is factive: when a subject S meets the conditions for knowing that *p* then it is a fact that *p*. What further platitudes might there be? It is common to assume that knowledge requires belief. Hence, it looks like we should also add the following principle to our minimal theory:

(BC)   Necessarily, if a subject S knows that *p* then S believes that *p*

However, the belief condition is not obvious to everyone (e.g. Prichard 1950). Moreover, some of the best arguments for this condition issue from certain reliabilist conceptions of knowledge. For that reason, I'm going to postpone discussion of the belief condition until §1.5. Let us ask instead: does knowledge require reliability?

Williamson (1994, Ch. 8, fn.5; 1997b, fn.5) has argued that reliability is a necessary condition on knowledge on any reasonable conception. Thus, even if one thinks that knowledge that *p* requires that one be sufficiently justified in believing *p*, then knowledge still requires reliability for to have sufficient justification for one's belief itself requires that one's belief be reliable. If Williamson is right then our minimal theory should sanction a reliability condition on knowledge. But what is reliability? And can this notion be minimally characterised?

Say that a belief is reliable if it is produced by a reliable process or method of belief-formation. Typically, a process or method is said to be reliable when it (generally) produces true beliefs and (generally) inhibits false beliefs. Goldman calls this the *global* reliability requirement (Goldman 1986, pp. 50-51). Alston (1993, pp. 528-9) takes this requirement to be a special case of the more general requirement that the reasons for holding one's belief should be 'truth-conducive'. The *basis*, i.e. the reasons for holding one's belief, or the method via which one has acquired one's belief, should 'probabilify' the belief, i.e. make it likely to be true (and unlikely to be false). Importantly, this requirement allows that one's belief can be formed on a good basis (or produced in the right sort of way) but can be

nonetheless false. Thus, this notion of reliability allows that one can have a justified but false belief. We can thus offer the following minimal condition:

(**GR**)   Necessarily, if S knows that *p* then S's belief that *p* is globally reliable.

Call this the global reliability condition, where we may simply define global reliability in Alston's more generic sense given above. Should we grant that **GR** is an axiom of epistemic minimalism?

Arguably, *yes*, but in doing so we must recognise that **GR** is unspecific in nature. We have yet to be told to what degree the *basis* for one's belief should probabilify this belief in order for this belief to count as globally reliable. Should the basis for one's belief simply make one's belief more likely to be true than false or should we set some more demanding threshold? It surely has to be conceded that our minimal theory is unable to supply a rule which allows us to say in every case whether a belief counts as globally reliable. Moreover, in certain cases one might allow that knowledge can be acquired via testimony from a person who is not generally reliable but is nonetheless believed to be so. Mr Grass, my informant, has his good days but more often than not he has his bad days. It just so happens that I catch him on a good day and he tells me that a South London gang are going to steal the Crown Jewels. He is right, and on the day that he told me this he provided me with very good evidence even though more often than not his information is misleading. It's not obvious in this case, that I do not know that the gang are going to steal the Crown Jewels.[20]

A second problem with **GR** is that it cannot be employed to explain why we intuitively lack knowledge in the standard Gettier cases. Take the following case: I believe that my brother is in Bristol because I see his car driving in the centre of the city and I catch sight of his wife in the passenger seat. My belief is justified as his car is highly distinctive and I have never known him let anyone else drive it apart from his wife. Moreover, my belief is true as (unknown to me) my brother has just arrived by taxi from the airport. (Indeed my brother's car has been stolen and his wife has been abducted.) Even so, my belief is both justified and true. Yet it intuitively falls short of knowledge. But why do we lack knowledge in such a

---

[20] It might be possible to resolve such problems, at least in part, by indexing **GR** to the method of belief-formation employed. If my method of belief formation is 'inferring from what Mr Grass says on any day of the week to believing what Mr Grass says is true' then this method is not globally reliable. However, if my method is 'inferring from what Mr Grass says on his good days to believing what Mr Grass says is true' then there is a sense in which I can gain knowledge that the gang will steal the Crown Jewels. The problem in this case emerges when we further demand that I must know or be able to discriminate Mr Grass's good days from his bad days. Suppose I cannot do so, might I still be said to have some form of *lucky* knowledge? We return to such issues in the next chapter.

case? My belief also seems to be globally reliable. The basis for my belief is a good basis. On other occasions, were I to employ the method of inferring from the presence of my brother's car and wife to the presence of my brother then this method would generally yield the true belief that my brother is in my locality (given all the other things I know). Since my belief is globally reliable, then one cannot employ **GR** to say why I lack knowledge that my brother is in Bristol. So, even if we allow that **GR** is a minimal axiom, we require some other principle to predict my ignorance in such a case. What form might such a principle take?

One common diagnosis of such Gettier cases runs as follows: a subject cannot be said to know in cases where it is only by chance that their justified true belief is true. Roughly, the thought here is that my justified true belief that my brother is in Bristol falls short of knowledge because it is *true by accident* (see Zagzebski 1994). This familiar diagnosis is seductive but open to misinterpretation (a theme which will be brought into greater relief in the next chapter). It all depends on what is meant by the expressions 'accidental true belief' and 'non-accidental true belief'.[21]

One natural explanation for our unwillingness to say that I know that my brother is in Bristol is given by the following principle

(**RC**)   Necessarily, if S knows that *p* then S's belief that *p*, or S's reasons or justification for their belief that *p*, are relevantly connected to the fact that *p*

(cf. the slightly different formulation given by Merricks 1995, p. 850). Say that a belief that *p* is non-accidentally true just in case this belief, or the reasons for this belief, bear the right sort of connection to the fact that *p*. Can **RC** provide a canonical explanation of why we lack knowledge in the standard Gettier cases?

My evidence for my belief that my brother is in Bristol is not in any way connected to the fact that he is in Bristol. My evidence depends on my seeing his car and wife in the passenger seat (plus certain background beliefs). But the fact that he is in Bristol is unrelated to my evidence. Hence, there is no relevant connection between my evidence for my belief and the fact believed. But there are a number of conspicuous problems with **RC**.

The first worry concerns whether and how we can give an explicit characterisation of the notion of 'relevant connection'. Is this connection causal, is it to be expressed using subjunctive conditionals, or can we best express this relation in some other way? Even if a

---

[21] Indeed, Unger (1968), is content to simply define knowledge as non-accidental true belief, but relies on an intuitive understanding of what counts as an non-accidental true belief.

neutral characterisation were available, there would nonetheless remain the problem of how to apply this principle in specific cases. What counts as a relevant connection will typically depend on specific features of the context in hand. Hence, to know whether or not to ascribe knowledge case by case will typically depend on a sensitivity to just those features.[22] One response to such worries is to leave the notion of relevant connection unspecified, such that we allow each substantive theory of knowledge to flesh out the principle **RC** according to constraints local to that theory. On this view, **RC** is simply a schema—one of Williams's highly formal remarks, which is available to every theory in virtue of its unspecific nature.[23]

A related and perhaps more pressing worry concerns the scope of principle **RC**. While this principle can be employed to correctly predict ignorance in the standard Gettier cases, as this principle is stated, it appears unable to predict why we lack knowledge in a range of more sophisticated cases. The most famous case of this sort concerns Goldman's case of Henry in barn-facade land (Goldman 1976). A structurally similar example runs as follows: on his travels, Jimmy comes to a place where *nearly* everything is not what it appears to be. Unluckily for Jimmy, he has arrived in fake-world. The fake objects in fake-world are so well constructed that just by looking Jimmy cannot tell that they are not real. Even so, Jimmy continues to form beliefs about his environment in fake-world. One particular feature of fake world is that there a lots of fake dogs hanging around. Jimmy forms the belief that there is a (real) dog crossing the road. As it turns out, it is a real dog and not one of the many fakes. His belief is nonetheless justified. He can cite good reasons for it, and indeed his method of belief formation is globally reliable—in the past, it has always yielded true beliefs about what is and what is not a dog. Moreover, his belief, and the reasons for his belief, are relevantly hooked up to the facts, so goes the thought, since the fact that there is a dog crossing the street is causing him to perceive that there is a dog crossing the street (given his disposition to form beliefs about objects in his environment). Thus, not only does Jimmy have a justified true

---

[22] Again, this is just another facet of the generality problem (see fn.11 above).

[23] One might have the further worry that whether one's belief is non-accidentally true (relevantly connected) may be a fact which is not reflectively accessible to the knowing subject. On certain internalist conceptions of knowledge, in order to know that *p* I must know that I know that *p*. Suppose I do know that *p*, and hence my belief (or the reasons for my belief) bear an appropriate relation to the facts. Can I thereby know, upon sufficient reflection *from the inside*, that my beliefs or reasons do indeed bear the right sort of relation to the facts? Take the famous case of the industrial chicken-sexers (see Brandom 1998). Surely the chicken-sexers can cite reasons for their belief that their first-order beliefs concerning the sex of chicks are relevantly connected—they can justify this second-order belief by adverting to the fact that they have not been sacked by the factory manager. They do not know *how* they know the sex of chicks, but they do know *that* they know. Thus, there are grounds to think that whether one is relevantly connected to the facts can be known (or at the very least justified) since one's belief that one is relevantly connected is well-supported by evidence which one can bring to mind. Like Brandom, I am inclined to think that examples like the chicken-sexer present no particular problem for the internalist (see §2.5 below)..

belief, his belief is relevantly connected to the facts. Just about everybody agrees that in such cases Jimmy lacks knowledge that he is seeing a dog. But why?

Perhaps the most popular and enduring diagnosis of cases of this sort is that Jimmy's belief is just too lucky to count as knowledge. He could so easily have gone wrong in forming the belief that a (real) dog was crossing the street as it might so easily have been a fake dog. Here the thought is that despite there being a causal connection between Jimmy's belief and the facts, this connection is not relevant since Jimmy's belief is just too accidental to count as knowledge. Fake-world is a knowledge-unfriendly world, a world where the risk of being in error is just too great to allow knowledge.[24] Call such Gettier cases *anti-luck* Gettier cases.

The possibility of anti-luck Gettier cases highlights the deficiencies of principle **RC**. On the one hand, there does seem to be a relevant connection between Jimmy's belief and the facts, since his belief is in some sense caused by the facts, and yet on the other hand, there does not seem to be the right sort of relevant connection, since his belief remains accidentally true. It is for this very reason that Goldman (1976, 1986) suggests that in order to correctly predict ignorance in anti-luck Gettier cases, we should replace the principle **RC** (and cognate principles) with a more specific reliability condition on knowledge, namely what he calls a *local* reliability condition which holds in addition to the global reliability condition **GR**.[25] In knowledge friendly environments (where it is not easy to make mistakes), one will be both locally and globally reliable, but in knowledge-unfriendly environments (where it is easy to make mistakes), one will simply be globally reliable. What exact form should such a local reliability condition take?

1.5 Local reliability: sensitivity, stability, safety and robustness

Perhaps the most enduring and popular local reliability condition on knowledge is what is called the *sensitivity* condition, which we may give as follows:

---

[24] Zagzebski (1994, p.66) thinks that we lack knowledge in such cases because an accident of good luck (hitting on the truth) is cancelled out by an accident of bad luck (being in a predicament where our methods of belief-formation would not generally yield true beliefs). This is not the only possible diagnosis. On a defeasibility analysis of knowledge, knowledge is undefeated justified true belief. Thus, Jimmy can be said to lack knowledge because there is relevant evidence which he does not possess, namely the true defeater *Jimmy is in fake-world*. Were Jimmy to become aware of this fact, he would be no longer justified in believing that he is seeing a dog. His justification is defeasible, and so he lacks knowledge (Harman 1973; Pollock 1986).

[25] In his (1967), where he defends a causal theory of knowledge, Goldman does not employ the principle **RC** as such, but rather the more specific condition 'S knows that p only if S's belief that p is caused by the fact that p'.

15

(**SEN**)   Necessarily, if S knows that *p*, then if *p* were false S would not believe that *p*.

Roughly, this principle says that to know that *p* one's belief in the actual world must be sensitive in the sense that in the nearest not-*p* worlds one does not believe that *p*. **SEN** is also a condition sanctioned by the nomic-causal theory of knowledge given by Armstrong (1973), the causal-reliability theory of Goldman (1976, 1986), and the Relevant Alternatives theory of Dretske (1970, 1971) and Nozick (1981).[26] These authors exploit **SEN** to predict why a subject lacks knowledge not only in the standard Gettier cases, but also in anti-luck Gettier cases. I lack knowledge that my brother is in Bristol, for were he not to be in Bristol I would nonetheless believe that he is in Bristol (given the evidence I have). Jimmy lacks knowledge that he is seeing a dog in fake-world, for were he not seeing a real dog but a fake dog he would nonetheless believe that he is seeing a dog.

Nozick (1981, pp. 22-4) famously noted that that the sensitivity condition is satisfied in certain anti-luck Gettier cases where one intuitively still lacks knowledge, namely the case of the assassinated dictator (due to Harman 1973, pp. 142-154), and his own case of the envatted brain who is 'fed' the belief that it is envatted. In the former case, Jill sees a news report that the dictator of her country is dead, but then later she misses just by chance the (false) report in which the original story is denied. Nozick argues that even though Jill's belief that the dictator has been assassinated is true and is indeed sensitive (were it false that he has been assassinated she would not believe that he was assassinated) it does not satisfy the stability condition. In cases close to the actual case, the dictator has indeed been assassinated but she would have heard the news reports denying the murder in these cases and so would not have believed that he is dead after all. Jill's sensitive true belief is unstable and thus cannot be knowledge. In the envatted brain case, one's true belief that one is a brain-in-a-vat is likewise sensitive (were you not a brain in vat you would not believe that you were) but unstable: the belief that you are envatted is easily lost in (some) nearby cases for in those cases the scientists no longer feed you that belief.

---

[26] Nozick's thumbnail account of the semantics for the subjunctive 'if *p* were true *q* would true' runs: 'the subjunctive is true (roughly) if and only if in all those worlds in which *p* holds true that are closest to the actual world, *q* is also true. (Examine those worlds in which *p* holds closest to the actual world, in see if *q* holds in all these.) Whether or not *q* is true in *p* worlds that are still farther away from the actual world is irrelevant to the truth of the subjunctive' (Nozick 1981, in Luper-Foy 1987b, p. 21) Note all page references to Nozick (1981) will refer to the excerpt which appears in Luper-Foy 1987b). A more specific formulation of **SEN** would index this principle to the method of belief formation, but I shall omit such niceties, important as they are, here (see Nozick 1981, pp. 2-4). A principle of sensitivity also features appears in the contextualist theory of knowledge given by DeRose (1995) as a principle which can be employed to specify the epistemic standards which may hold in particular contexts.

Nozick (1981, p. 23) therefore proposed an additional local reliability condition on knowledge, which we may call the *stability* condition, given as follows:

(STA)   Necessarily, if S knows that *p* then were *p* true then *S* would believe that *p*.

This principle says that to know that *p* one's belief in the actual world must be stable in the sense that in the nearest *p* worlds one believes that *p*.

For Nozick, a stable and sensitive (true) belief just is knowledge. Given that from the perspective of epistemological minimalism we have jettisoned the ambition to provide a constitutive definition of knowledge then we are only here interested in the claim that knowledge that *p* requires that one's belief that *p* be both sensitive and stable. Can the minimal theory enjoin this conjunctive condition on knowledge? Such a condition is open to objection on a number of counts, particularly the requirement of sensitivity.[27]

It's a familiar worry that the sensitivity condition concedes too much ground to the sceptic. Take the hypothesis that I am not a brain in vat. Given SEN, this hypothesis is not known to be true: were it not the case that I am not a brain in a vat (that is, were it the case that I am a brain in a vat) I would nonetheless believe that I was not a brain in a vat—my belief is thus insensitive, and I lack knowledge. In Nozick's phrase, the world in which I am envatted and the world in which I am not are 'doxically identical'. The sceptical impact of such a concession is mitigated by denying both that knowledge transmits 'forwards' and ignorance transmits 'backwards' across known entailments: while one cannot be sensitive to the fact that I am not a brain in a vat this does not rule out the possibility of being sensitive to the fact that I have two hands. Though influential, such an analysis is far from convincing. For one thing, the transmission of knowledge is a highly intuitive principle; for another, it's not clear that scepticism has been defused in any substantial way.[28] The fact that such a local reliability condition has proved to be so problematic and controversial makes it strictly unavailable to the epistemic minimalist. Is there another less controversial local reliability condition?

---

[27] See Williamson (2000, pp. 147-56) for an excellent evaluation of the sensitivity condition

[28] Nozick quite clearly takes the business of the anti-sceptic to consist in showing that ordinary knowledge is *possible* (see p. 41). But such a task is really only a necessary and not a sufficient response to the sceptic. For all that Nozick has said, the actual world or some nearby world might be a world in which all our beliefs are false. If that were so, we would lack knowledge. So for all we know, we do lack knowledge, for we cannot tell that we are reliably tracking the truth. Thus a response to the sceptic must show us (give us a guarantee) that we do indeed have knowledge—it must give us knowledge that we have knowledge and not merely knowledge that knowledge is possible. (See Stroud (1994) for the best discussion of so-called meta-epistemological scepticism.)

Williamson (1994, 2000), Sosa (1996, 1999, 2000), and Sainsbury (1997) have all suggested that knowledge requires *safety* not sensitivity, though they differ somewhat as to how this notion of safety is to be best expressed.[29] Williamson defines safety as follows:

in a case α one is safe from error in believing that C obtains if and only if there is no case close to α in which one falsely believes that C obtains (2000, pp. 126-7).

(where the degree of closeness will not only depend on the particular method of belief formation being employed but also on features of the context). Williamson thus offers the following safety condition on knowledge:

(**SAF**)   Necessarily, if S knows that *p* then it is not a close possibility that one falsely believes that *p*.

Williamson takes it that a belief is reliable just in case it is safe, while Sosa (1996) seems to think that a reliable belief is both safe and stable. Sainsbury thinks that there may indeed be more than one safety requirement—something I will come to in a moment.[30]

The safety and sensitivity conditions issue in radically different conceptions of knowledge. For one thing, safety, unlike sensitivity, entails no immediate concession to the sceptic. There is thus no requirement that one must reject the closure principle given the safety condition— one simply does not concede that one does not know that the sceptical hypothesis is true.[31] It looks like the minimal theory can and should reject sensitivity in favour of safety as a condition on knowledge. But should the minimalist also endorse the stability requirement **STA**?

Sainsbury (1997), however, offers good reasons why the stability condition appears to be too strong. Here's his counter-example:

---

[29] They each bracket consideration of what the sufficient conditions of knowledge might be.

[30] A key antecedent of the view that knowledge requires safety not sensitivity is Luper-Foy (1987a, p. 234). Sosa (1999) gives the safety requirement in terms of the counterfactual conditional: If S were to believe that *p* then *p* would be true. Since counterfactual conditionals do not in general contrapose then safety and sensitivity are non-equivalent conditions, though the two conditions are, he notes, easily confused. Arguably, it is better to state the safety condition as Williamson does, then we at least avoid the problem of just which semantics for subjunctives is to be favoured.

[31] Safety-reliabilism thus opens up interesting possibilities with respect to the sceptical threat (see Sainsbury 1997, Sosa 1999, and Williamson 2000, Ch. 8). On a simple version of safety-reliabilism, it is not an easy possibility that I am a brain in vat and so *a fortiori* it is not an easy possibility for my belief that I am not a brain in a vat to be false. At the very least, this leaves open the possibility that I can know that I am not a brain in a vat. On these grounds alone, safety-reliabilism is an attractive view. Of course, more sophisticated versions of safety reliabilism owe us an account of just why it is not an easy possibility that I am envatted. (See Sainsbury 1997, and Williamson 2000, Ch.8.)

I believe that Mary is married because I notice her wedding ring. She is married but she hardly ever wears the ring and indeed didn't intend to wear it on this occasion (she left it on by accident). I have no curiosity about her marital status, so if I had not noticed the ring I would have formed no belief about whether or not she was married. Arguably this falsifies [STA], so I don't track the truth. Yet, arguably, I couldn't easily have been wrong (given the uniformly prevailing convention in my culture that only married persons wear such a ring) (Sainsbury 1997, p. 909).

For Sainsbury, such cases represent cases of what we may call *lucky* knowledge—knowledge that we might easily not have had. He indeed knows that Mary is married, despite it being a matter of luck that she was wearing a ring on this occasion and so it being a matter of luck that he thereby formed the (correct) belief that she is married.[32] However, if stability plays no part in our understanding of reliability and non-accidental true belief then what are we to say of Nozick's original motivations for introducing the stability condition? Though Sainsbury does not explicitly discuss this problem, it seems important to address it.

Do we have lucky knowledge (knowledge we might easily not have had) in the cases like the assassination case and the case of the envatted brain who is being whimsically fed the belief that it is a brain in a vat? If we don't intuitively have knowledge in these two cases then what can possibly explain our ignorance in these anti-luck Gettier cases once we have dispensed with the stability condition on knowledge? Note that even if we express the stability requirement without employing counterfactual conditionals as follows:

(**STA***) One's belief that *p* is *stable* just in case it is not an easy possibility both that *p* is true and one does not believe that *p*.

the problem still remains. In Nozick's two cases, as well as Sainsbury's married-Mary case, it is an easy possibility that *p* is both true and not believed. But if we recognise the possibility of lucky knowledge in the married-Mary case, this new version of the stability condition **STA*** cannot be employed to rule out knowledge in Nozick's anti-luck cases either. That said, there are important differences between the married-Mary case and Nozick's two cases

---

[32] Even if Nozick's stability condition is relativised to methods of belief formation such that a belief that *p* is stable just in case were *p* true, and *S* were to use a method M to arrive at a belief whether (or not) *p*, then *S* would believe, via M, that *p* (Nozick 1981, pp. 25-31), this does not rescue the stability condition. The belief that Mary is married is acquired using the method of inferring from the appearance of a wedding ring on a woman's finger to the belief that the woman is married. Employing the same method in close cases produces no belief that Mary is married (she does not wear the ring in close cases and its absence is only sufficient for me to withhold belief about Mary's marital status). So, even the relativised condition fails even though intuitively we should say that knowledge that Mary is married is not undermined—such knowledge is lucky, as Sainsbury

and many people are indeed likely to feel that knowledge is indeed absent in Nozick's two cases. But exactly why?

Notice that in the assassination case, had the second news report (in which the assassination is vigorously denied) been seen then Jill would, presumably, not only lose her belief that the dictator had been assassinated she would also gain the belief that he had not been assassinated.[33] In the close case, she thus forms a false belief. Likewise in the case of the envatted brain. As the example is set up, it's an easy possibility that one would not only lose one's belief that one is envatted, but an easy possibility that one gains the belief that one is not envatted. Again, in the close case, one forms a false belief. What else could possibly explain our intuition that we lack knowledge in this case? (I shall return to this crucial question in a moment.) In the married-Mary case the matter is very different. In close cases, while one would lose one's belief that Mary is married, there is no likelihood that one would gain the belief that she is not married. All the evidence one has would tend to make one refrain from forming a belief—after all the prevailing conventions are such that not all married women wear wedding rings.

The observations just given suggest that the safety requirement is not the only condition on knowledge—we need a further condition which explains our ignorance in both the assassination case and the case of the envatted brain. Funnily enough, Williamson fails to acknowledge this, because Williamson ties reliability much too closely to safety (2000, pp. 123-6). Arguably, for one's belief that C obtains to be reliable requires not only safety, in that that it is not an easy possibility that one believes that C obtains when this condition does not obtain, as Williamson specifies, but also, what we may term, *robustness*, in that is not an easy possibility that one believes that C does not obtain when it does obtain. This is to say that there's a sense in which a belief can be safe but not reliable since it was produced by a process which yields beliefs which are not robust.[34] One's belief that *p* is not robust when there is some close case in which *p* is true and one acquires the belief that not-p. Hence, we define give the requirement of *robustness* as follows:

---

highlights. Thus it looks like it is mistaken to think as Sosa (1996) does that non-accidental true beliefs are both safe and stable. That said, Sosa (2000) is less convinced of the requirement of stability than in his (1996).

[33] Nozick's presentation is unspecific in that he intends his example to cover both scenarios. He says, 'everyone who encounters the denial believes it (or does not know what to believe and so suspends judgement)' (1981, p. 23). One's intuitions can only be trusted once one recognises the key difference between losing one's belief that *p* via employing a method M (listening to news reports) and gaining a false belief that *p* via the same method.

[34] Sainsbury (1997, pp. 908-9), unlike Williamson, recognises that the reliability conditional 'if one knows then one could not easily have been mistaken' to sanction both safety and robustness (though Sainsbury tends to speak of these cases as being two species of safety—but the difference between the text and Sainsbury's characterisation is merely terminological.)

(ROB) Necessarily, if S knows that p then it is not an easy possibility both that p is true and one believes that *not-p*.[35]

So, a belief that p is reliable just in case there is no danger that one will *overshoot*, as it were, and form the same belief when not-p and no danger that one will *undershoot*, as it were, and form the belief that not-p when p.

But what is the relationship between robustness and stability? It depends on the relationship between believing *not-p* and not believing p. Under favourable conditions (where a subject is (ideally) rational) if a subject S believes that *not-p* then S will not believe that p, but not vice versa. Stability entails robustness at least where this doxastic inference is valid.[36] However, robustness does not entail stability. That ought to be no great surprise. It can be an easy possibility that my belief that p be lost but not an easy possibility that I gain the belief that *not-p* (where p remains true in all close cases). The married-Mary case is just one such example.

Thus, it looks as if knowledge not only requires global reliability but also local reliability, at least so long as we say that a belief is locally reliable just in case it is both safe and robust. But what about the belief condition on knowledge?

1.5 The belief that p condition and the non-belief that not-p condition

Suppose we simply take a belief that p to be the mental state (or propositional attitude) of taking p to be true.[37] Should we then accept the following principle?

---

[35] Note that Nozick (1981, p. 25) augments his stability requirement to include the requirement of robustness when he offers the following counterfactual condition on knowledge: if p were true then S would believe that p *and* it's not the case that S would believe that *not-p*. But Nozick merely introduces the requirement of robustness to ensure that S does not hold contradictory beliefs not to ensure that one does not form false beliefs in close cases.

[36] This inference is an instance of Hintikka's doxastic rule BC (Hintikka 1962). We can see that stability entails robustness as follows: Assume that one's belief that p is stable. Hence, if one believes that p in a case α then in a case β it is not the case that both one does not believe that p and p is true (where β is close to α). Assume for reductio that one's belief that p is not robust. Thus, one believes that p in a case α but in β one both believes that *not-p* and p is true (where α is close to β). From these two assumptions we can infer that in β it is not the case that both one does not believe that p and p is true. Since from the second assumption p is true in β, it follows by *modus ponendo tollens* and double-negation elimination that one believes that p in β. Since it also follows from the second assumption that one believes that *not-p* in β, then given Hintikka's rule BC, we can infer that one does not believe that p in β. Contradiction. Reject the assumption that one's belief is not robust and an application of conditional proof shows that stability entails robustness.

[37] Where 'p' schematises whatever one's favourite kind of primary truth-bearer happens to be. One key feature of belief is that it be governed by the teleological norm: in believing p, aim (at the very least) to form a belief that is true. Whatever else we may have to say about belief may well be far from uncontroversial.

(BC)   Necessarily, if a subject S knows that *p* then S believes that *p*

It is certainly controversial whether knowledge entails belief.[38] If one gives up the presupposition that knowledge can be analysed as a conjunction of truth, belief, and other more elusive features then one is at liberty to reject the belief condition on knowledge.[39] As it turns out, there are some bad arguments for giving up the belief condition and no particularly good arguments (beyond mere intuition) for retaining the belief condition.

In ordinary language, one may indeed hear people say: 'John no longer believes that Mary is unfaithful, he now knows it'. Taken at face value, this suggests that we give up the belief condition since it suggests that knowing and believing are incompatible states. But one might easily read this sentence as saying that 'John no longer *merely* believes that Mary is unfaithful, he now knows it'; in which case one can allow that the belief condition remains intact. A slightly more persuasive argument is given by Lewis (1996, p. 550). Lewis argues that knowledge does not entail belief using the (well-worn) example of the nervous school-child 'who knows the answer but has no confidence that he has it right, and so does not believe that he knows'. But such an example is unconvincing. Lewis seems to assume that all beliefs are reflectively accessible to the believer and indeed that to believe that *p* entails that one is confident that one is right about *p*. But one can surely believe something without under certain circumstances manifesting any confidence in the truth of what one believes. Nor indeed need our beliefs be reflectively accessible to us: as I walk along the cliff I believe that I am safe from falling but were I to reflect on my belief I may no longer believe that I was safe.

Perhaps the best consideration for retaining the belief condition comes from reflecting that knowing that *p* does seem to involve a commitment that *p*: to know *p* entails taking *p* to be to be true. If we grant that a belief that *p* is to be minimally defined as *a taking p to be true*, it follows that knowledge entails belief (see Zagzebski 1999 p. 93). But beyond the mere intuition that knowledge entails belief, can we construct an argument for this entailment?

Here's one such argument: assume that knowledge entails stability. That is, if one knows that *p* in a case α then there is no case β close to α in which *p* is true and one does not believe

---

[38] Prichard (1950) took the strong view that knowledge excludes believing.

[39] Rejecting the belief condition does not of course entail that every case of knowing excludes believing. Williamson rejects a componential analysis of knowledge but does not reject the belief condition. Instead, Williamson construes *mere* believing as a botched kind of knowing (Williamson 2000, pp. 41-8).

that p. Assume that one knows that p in α. Since the *is close to* relation is a reflexive relation (α is easily possible relative to α) then in α it is not the case that both p is true and one does not believe that p. But given factivity, p is true in α. Thus by *modus ponendo tollens* and double negation elimination, one believes that p in α. A step of conditional proof secures the required entailment. For those who accept the stability condition **STA,** this proof ought to be compelling. But we saw above that Sainsbury gave us good reason to reject **STA** given the possibility of lucky knowledge. But might knowledge that p entail something weaker than belief?

Assume that knowledge requires robustness. That is, if one knows that p in a case α then there is no case β close to α in which p is true and one believes that *not-p*. Assume that one knows that p in α. Since *is close to* is a reflexive relation, then in α it is not the case that both p is true and one believes that *not-p*. But given factivity, p is true in α. Thus by *modus ponendo tollens* and double negation elimination, one does not believe that *not-p* in α. A step of conditional proof secures the entailment from knowledge that p to the condition of not believing that *not-p*, a condition weaker than belief that p. Since we have as yet found no grounds to doubt the requirement of robustness, then it looks as if the minimal theory of knowledge should at least sanction the following condition:

(**NBN**)   Necessarily, if a subject S knows that p then S does not believe that *not-p*.

But can we retain the belief condition on grounds of intuition alone? Indeed, we know that this condition is controversial, so it's not obvious that it can feature as minimal axiom. Provisionally at least, I suggest that we retain the belief condition **BC**. As it turns out, in the next chapter, there will scope to adjust our minimal commitments, including the belief condition.

1.6 The minimal axioms

To recap, though we have found that no minimal definition of knowledge is available to the minimalist, it nonetheless seems that from the perspective of minimal epistemology one should at least provisionally accept the following list of axioms:

(TC)	Necessarily, if a subject S knows that *p* then *p* is true
(BC)	Necessarily, if a subject S knows that *p* (via method M) then S believes that *p* (via method M)
(NBN)	Necessarily, if S knows that *p* (via M) then S does not believe (via M) that not-*p*
(GR)	Necessarily, if S knows that *p* then S's belief that *p* is globally reliable.
(SAF)	Necessarily, if S knows that *p* (via M) then it is not an easy possibility both that *p* is false and one believes that *p* (via M).
(ROB)	Necessarily, if S knows that *p* (via M) then it is not an easy possibility both that *p* is true and one believes that *not-p* (via M).

There are three things to note. Firstly, we have retained the general reliability condition **GR**, but dispensed with the relevant connection condition **RC** in favour of the two conditions **SAF** and **ROB**. Secondly, where appropriate, each of these principles has been indexed to a method of belief formation. In general, it is advisable to do this, in order to avoid superficial counterexamples (see Williamson 2000, p. 128). Take the safety condition. I might know that *p* in a case α, but my method M of gaining this knowledge might be highly discriminatory such that it allows there is a close case β in which not-*p*. Using the method M, I will not form a belief that *p* in the close case β, but I may well form a belief that *p* in the β case via a method which is significantly less discriminatory than M. Lastly, let me stress that this list is decidedly provisional. As we shall in the next chapter, there is not only scope to question whether some of these principles really should feature in our minimal theory, we shall also find that there is a further key epistemic principle which ought to be acceptable to all partisans.

CHAPTER TWO

ARE WE ALL EXTERNALISTS NOW?[1]

In the previous chapter, the aim was to lay the foundation for a minimal theory of knowledge. The result was no more than a set of basic platitudes which, *prima facie* at least, state some of the necessary conditions on knowledge. In this chapter, the overall aim will be to test whether our minimal axioms are genuinely minimal, discard those that conflict with any further minimal intuitions we may have concerning knowledge, and put the remaining principles to work by seeing if we can derive some fruitful and perhaps controversial theorems from them. As it turns out, though epistemic minimalism is all but based on platitudes, it nonetheless promises to be able to settle certain pressing and long-standing issues in epistemology. In particular, our minimal theory of knowledge proves to be sufficiently rich to refute two key internalist theses, namely: that knowledge iterates (roughly, if one knows and one has properly considered whether one knows then one knows that one knows); and, that the mental realm is transparent to the knowing subject (roughly, if one is in a certain mental state and one has properly considered whether one is in that state then one knows that one is in that state).

In the previous chapter, following Williamson (1994, 1997b), it was more or less taken for granted that knowledge minimally requires local reliability. (Unlike Williamson, however, it was argued that a reliable belief or method, is not merely safe, it is also robust.) Williamson (1992a, 1994, 2000) has employed the safety requirement on knowledge to serve as the

---

[1] Most of the main ideas in this chapter were presented at a symposium on Tim Williamson's book *Knowledge and Its Limits* given at the conference *The Limits of Warrant*, University of Waterloo, Waterloo, Canada, May 18-20th, 2001. Many thanks to Tim Williamson for his very helpful reply. Elements of this chapter have also been incorporated into a review of *Knowledge and Its Limits*, *Times Higher Educational Supplement*, April 12th, 2002.

foundation for a general model of what he terms *inexact* knowledge. Since this model is derivable from our minimal axioms, plus other apparently acceptable principles, then it ought to feature in our minimal theory of knowledge. In §2.1-§2.4, Williamson's model of inexact knowledge is evaluated in detail and its minimalist credentials are in fact found wanting. One main feature of the discussion will be that it is simply mistaken to think that knowledge requires a local reliability condition.

In §2.1, we look at how Williamson uses his model of inexact knowledge, and particular his *margin for error principles*, to undermine the principles of iterativity and transparency mentioned above. These margin for error principles tell us that knowledge that *p* requires that it is not an easy possibility that *not-p*. In §2.2, I show how Williamson derives these principles from the thesis that knowledge requires safety, plus a key doxastic principle (roughly, the principle that there is limited discrimination in the belief-forming process). In §2.3, the pedigree of this doxastic principle is questioned on two counts: this principle does not comport well with Williamson's *knowledge-first* methodology, and it is not obvious that this principle is necessary or knowable *a priori*. This latter worry impinges on whether one can exploit Williamson's margin for error principles in support of a realist conception of truth—this is an issue which I return to in §3.8.

In §2.4, I assess whether it is right to say that knowledge requires a (local) reliability condition—either in the guise of Williamson's safety requirement or in the guise of something like Nozick's more familiar tracking conditions. It is found that the phenomenon of lucky knowledge—knowledge we might easily not have had—requires us to reject *all* forms of the local reliability requirement. Hetherington (1998) has further argued that a subject has lucky knowledge in *all* Gettier cases where this subject has a justified true belief. It is argued that while Hetherington is right to posit lucky knowledge in the well-known Henry-barn-facade case and the Jill-assassinated-dictator case, he is wrong to posit lucky knowledge in the Smith-Nogot-Ford existential generalisation case and similar cases. In §2.5, a case is made for the instatement of a causal version of the relevant connection condition on knowledge encountered in Chapter One in place of any local reliability condition. This condition allows us to distinguish Gettier cases where we have lucky knowledge from Gettier cases where we lack knowledge. Some of the problems and prospects for a causal theory of knowledge are evaluated and a case is made for rehabilitating what has long been thought to be a defunct view.

Hetherington has also argued that cases of lucky knowledge show that we should reject the project of modal epistemology altogether. Very roughly, Hetherington takes the lesson of

lucky knowledge to be that whether one knows does not depend in any way on what conditions obtain in nearby possible worlds. While Hetherington is right to reject local reliability as a condition on knowledge, he is wrong to reject the ambitions of modal epistemology. Modal epistemology and lucky knowledge are compatible if one suitably weakens the modal conditions for knowledge. This insight informs the minimal theory of knowledge developed in §2.6. This theory utilises what I term *minimal margin for error principles*, principles which tell us that knowledge that *p* requires that it is not an easy possibility that one knows that *not-p*. Given these basic modal principles, we can nonetheless still show that the internalist principles of iterativity and transparency given above both fail. So, from axioms which are at least *prima facie* acceptable to everyone, it follows that we are all externalists now.

2.1 Margins for error, luminosity, and knowing that one knows

Williamson (1992a, 1994, Ch.8, and 2000, Ch.4 and Ch.5) has employed the safety requirement on knowledge to ground a model of what he terms *inexact knowledge*. It is surely an uncontroversial datum that the methods via which we acquire knowledge are not perfectly discriminatory. Each time we employ these methods they do not always yield true beliefs (or inhibit false beliefs). Consequently, much of our everyday knowledge is inexact. By looking, I do not know the exact height of the person I see loitering in the garden even though I know that they are not exactly seven feet tall and that they are not exactly five feet tall. Williamson claims that cases of inexact knowledge are governed by principles requiring knowledge to leave room for a *margin for error*—a cushion, as it were, outside of which we are safe from making mistakes. Such principles typically take the following form:

> (ME) For all cases α and β, if β is close to α and in α one knows that a condition C obtains then C obtains in β.

(where α and β range over actual and counterfactual cases). ME effectively says that if I know that C obtains then it is not an easy (or close) possibility that C does not obtain.[2] What counts as close? Closeness is more or less fixed by the method one employs in gaining

---

[2] Or we can express the matter thus: "'A' is true in all cases similar to cases in which 'it is known that A' is true' (Williamson 1994, p. 591). The term 'easy possibility' is taken from Sainsbury (1997).

knowledge (though Williamson concedes that contextual factors may also play a part). He adds:

> even if in α one believes that C obtains and is safe from error in doing so, it does not follow that C obtains in every case close to α, for there may be cases close to α in which C does not obtain and one does not believe that it obtains [...] if whether one believes is sufficiently sensitive to whether C obtains. For example, one may be safe from error in believing that the child is not falling even though she is not safe from falling, if one is in a good position to see her but not to help her (2000, pp. 126-7).

Thus knowledge and falsity (but not false belief) may indeed be close—at least if one's method of acquiring knowledge is sufficiently discriminatory. Less discriminatory methods require a greater margin for error. Where our methods are perfectly discriminatory, when they are *absolutely reliable*, then knowledge requires no margin for error at all.

At first blush, such a model seems to be uncontroversial enough to feature in our minimal theory of knowledge. Indeed, Williamson derives **ME** from the safety requirement on knowledge plus other principles which he takes to be uncontroversial. However, on the basis of such principles as **ME**, Williamson constructs both a novel and powerful case against epistemological internalism and the Cartesian conception of the mental. So while the axioms of his theory appear benign enough, some of the theorems of his theory are decidedly controversial. To demonstrate this, we first need to define some terms. Say that a certain condition C is *luminous* just in case the following principle holds:

(PK) If C obtains then one is in a position to know that C obtains.

Among the chief candidate conditions for being luminous are what we may call the core Cartesian states of being in pain, feeling cold, feeling hot, and so on. One might not want to include on this list states which are rather more complex in character, such as feeling jealous, feeling guilt, or feeling schadenfreude.[3] Even so, one might want to include states which consist of having a particular propositional attitude. If I'm in the state of believing that there is milk in the fridge, then I'm in a position to know that I have this attitude. Likewise for hoping, supposing, desiring, denying, wondering, and so on.

Being in a position to know (unlike being physically and psychologically capable of knowing) is factive. Where one is in a position to know *p*, the fact that *p* is readily available

---

[3] There seem to be states intermediate between the core Cartesian states and these more complex states, for example, feeling stressed, feeling tired, feeling happy, or feeling drunk.

to be known—one has merely to take a proper look, as it were, and knowledge results (Williamson 2000, p. 95). Given the following plausible principle:

(P) If one is in a position to know that C obtains and one has properly considered whether C obtains then one will know that C obtains

one can derive the weak principle of omniscience:

(K) If C obtains and one has properly considered whether C obtains then one knows that C obtains.

A stronger thesis of omniscience omits to add the proviso of proper consideration.[4] Such a principle is wholly implausible and is not the target of Williamson's analysis.

Epistemological internalism with respect to the mental typically entails K.[5, 6] Such a view not only subsumes the view that that the bearer of the core Cartesian state is in a much better epistemic position than anyone else with respect to knowing whether they are in that state (what is usually called 'privileged access'), it entails that one's sincere, first-person, present-tense, claims about one's core mental states are both infallible (immune to error), and incorrigible (immune to correction).[7]

---

[4] There is a presumption that proper consideration includes the aim to know. Hence, in properly considering whether one knows one actively seeks to form a belief whether one knows.

[5] Shoemaker (1988, 1994) takes K, when applied to mental states, to express the fact that such states are *self-intimating*: 'To say that a mental state is necessarily self-intimating means that it follows from someone's having the state that the person is aware of having it, or, on a weaker version of the notion, that the person would be aware of having the state if she or he considered the matter' (*ibid.*, p. 396). Guttenplan (1994, p. 291), agrees with this formulation, but adds that 'it should be noted that [self-intimation] does not by itself require that we are either infallible or incorrigible about our pains or other mental contents'. The argument in the text which follows conclusively demonstrates that self-intimation does require that our methods of acquiring knowledge that we are in a certain mental state be infallible.

[6] We must take care to distinguish epistemological internalism with respect to the mental from *semantic* externalism. If content is to be (partly) individuated by external environmental factors, then it is plausible to think that the attitude a thinker takes with respect to a particular content is likewise to be (partly) dependent on external factors. The challenge then is that if mental states are, crudely speaking, not in the head, then it at least follows that we are not in a position to know whether we are in a particular mental state merely by some *internal* process of reflection or introspection. One can then go one of two ways: allow that reflection or introspection can give us knowledge of our environment, such that a knowledge of our mental life is achievable though reflection, or, disallow that introspection or reflection can yield knowledge of our environment, and thus reject the luminosity of our mental states in almost every case.

[7] Such a conception of the mental can be caricatured as follows: the mind is just like a halogen lit, whitewashed room, whose contents are fully illuminated. One may not be attending to a particular content at any one time (I have bad toothache but someone tells a very funny joke and I forget momentarily that I am in pain) but a quick inspection of the room will be able to yield knowledge of its contents. Contrast that with the opposing caricature: the mind is a Dutch-barn (open to the elements on one side); from the ceiling hangs a 60-watt bulb which illuminates some of the central contents but is too dim to illuminate much beyond that. Just what goes on in the barn, is not always readily available to be known. (I have a torch, but my therapist sometimes has a much brighter one.)

There is also, of course, a long-standing internalist tradition which holds that certain epistemic conditions, in particular the condition of knowing that p, satisfy both **PK** and **P** and hence **K**.[8] Where C is the condition that one knows that p, then we can derive a version of the KK-principle as follows:

> (**KK**) If one knows p and one has properly considered whether one knows p then one knows that one knows p.

Internalism with respect to knowledge typically entails **KK**.[9] (Familiarly, it does not entail an analogue of the S5 axiom since even though the condition of knowing p may be luminous it does not necessarily follow that the condition of not knowing that p is luminous.[10])

Williamson (1996a, 2000, pp. 97-8) offers an important argument which shows that principles like **ME** rule out the possibility of there being any luminous mental states, which is to say, in Williamson's vivid phrase, that we have no 'cognitive home'.[11] He also offers a related though rather more complex argument which shows that **ME** is incompatible with **KK** (see his 2000, Ch.5 for the most detailed version of this argument).[12] Let's try to capture

---

[8] Advocates of this tradition include Prichard (1950, p. 86); Chisholm (1966, p. 117); Ginet (1970); Bonjour (1985, 1992).

[9] Importantly, **KK** as stated, is not a sufficient condition of internalism. A more specific statement of **KK** would involve indexing the second-order knowledge claim to a basis (or method of belief-formation). Internalists typically claim that one can gain second-order knowledge on the basis of reflection or introspection. Thus indexed, the principle constitutes the defining thesis of internalism (with respect to knowledge).

[10] Suppose one does not know that p. Even if one has properly considered whether one knows that p, it will not in general follow that one knows that one does not know that p. Such an entailment fails since under less than ideal evidential conditions there are standard counterexamples: the lighting is bad and I form the belief that your car is beige; in fact it is white and so I don't know that it is white; however, I don't form a belief about my ignorance (for one thing I'm not aware that the lighting is bad) and so I don't know that I am ignorant (*cf.* Williamson 1994, p.17).

[11] One must take care not to overstate Williamson's position. He does not wish to say that in every case of being in a core Cartesian state one will fail to know that one is in that state after due consideration. Rather, he merely wishes to say that our methods of knowing whether we are in a mental state are not absolutely reliable, such that one *can* be in a mental state and not know that one is in that state. No cognitive home, means no guarantee that the contents of your mental life are readily available to be known by you in every case.

[12] The anti-**KK** arguments given in Williamson (1992a), (1994), and (2000) each differ slightly though they all employ a version of the closure principle for knowledge. As Williamson (1994, pp. 227-8) notes, this principle is strictly superfluous to the result. The closure version of the argument effectively employs a contraposed margin for error principle which the subject (by reflection on his methods of belief-formation) knows to be true. A simpler version of this argument runs as follows, (where 'K' abbreviates 'It is known that'):

(1) $K[K[\text{The tree is not i inches tall}] \rightarrow \text{The tree is not i+1 inches tall}]$ Known margin for error
(2) $Kp \rightarrow KKp$ **KK**
(3) $K[\text{The tree is not 0 inches tall}]$ Given
(4) $KK[\text{The tree is not 0 inches tall}]$ 3, given **KK**
(5) $KK[\text{If the tree is not i inches tall}]$ then $K[\text{The tree is not i+1 inches tall}]$ from 1, given distributivity
(6) $K[\text{The tree is not 1 inch tall}]$ 5, 4, instantiation, MP

by repeating this inference pattern 1000 times a paradox results, for the tree, let us say, is in fact 1000 inches tall, and since knowledge is factive we have a contradiction. Upshot: reject **KK**.

the import of both these arguments by using a hybrid argument where we can think of the condition C as both standing for the core Cartesian conditions (feeling hot, being in pain, and so on), as well as the epistemic condition of knowing *p* (where one can think of *p* as abbreviating the proposition that a certain tree is less than ten feet high, for example). The margin for error principle **ME** entails:

(1) $(\forall \alpha)(\forall \beta)$ If in α one knows that C obtains, and β is close to α, then C obtains in β

(where α and β range over both actual and counterfactual cases). Suppose that the condition C is indeed luminous to the subject and that the subject is properly considering in every case whether C obtains then :

(2) $(\forall \alpha)$ If C obtains in α then in α one knows that C obtains

Given the transitivity of the conditional, it is easily shown that from (1) and (2) we can derive the unpalatable:

(3) $(\forall \alpha)(\forall \beta)$ If in α one knows that C obtains then in β one knows that C obtains

(where β is close to α). This principle is absurd, for if the condition C is known to hold in one case it follows that it is known to hold in all cases. The Cartesian conception of the mental is straightforwardly undermined by principles like **ME**. Likewise, take the epistemic condition of knowing that a certain tree is less than ten feet high. Given that there are cases where one knows that one knows that the tree is less than ten feet high (when it is one foot high, for example), then from (3) by successive applications of universal instantiation and *modus ponens* one can derive the absurd result that there are cases where one knows that one knows that the tree is less than ten feet high when in fact the tree is greater than ten feet high. Upshot: knowledge that *p* is not a luminous condition and so when one knows and has properly considered whether one knows it does not follow that one knows that one knows. Internalism with respect to knowledge in general (and not just self-knowledge) is straightforwardly ruled out by **ME**.

It's important to recognise that Williamson's results are novel. Indeed, they are not aimed at knocking down straw men. There are plenty of extant and sophisticated internalists with

respect to knowledge and with respect to self-knowledge.[13] A fairly plausible looking principle **ME** readily undermines both forms of internalism. While I have no worry with the above argument *per se*, there are doubts about the grounds for accepting ME. Let us see.

2.2 From safety to margins for error

For Williamson, **ME** is well-motivated because he takes it to be derivable from the safety requirement on knowledge—though he does explicitly note that the route is not exactly direct. He says that

> If we combine the safety requirement on knowledge with limited discrimination in the belief-forming process and some plausible background assumptions, then we can [verify **ME** and] deduce failures of luminosity (2000, p. 127).

In more detail, Williamson reasons more or less as follows: suppose we reformulate the safety requirement as follows:

> (**SAF**) For all cases α and β, if β is close to α and in α one knows that C obtains, then in β it is not the case that: (one believes that C obtains and C does not obtain).

Now suppose we also have the following doxastic principle, where c is some small positive real number, and where the obtaining of condition C obtains depends only the value $v(\alpha)$ taken by some parameter v in α, and where v takes non-negative real numbers as values:

> (**B**) For all cases α and non-negative real numbers u, if $|u-v(\alpha)| < c$ and in α one believes that C obtains then, for some case β close to α, $v(\beta) = u$ and in β one believes that C obtains

Informally, Williamson glosses this import of this principle as follows:

> if one has the belief, then one could easily still have had it if the parameter had taken a slightly different value [...] If one believes that the tree is at most fifty feet high,

---

[13] Epistemological internalism is more often than not defended with respect to justification (e.g. Lehrer 2000). Weaker margin for principles are available which serve to undermine the thesis that warrant (or justified belief) is iterative (see Williamson 1994, pp.244-7). See Feldman (1981) for some independent criticisms of internalism with respect to knowledge.

then one could easily still have believed that if the tree had been an inch higher, but not if it had been one hundred feet higher (p. 127).[14]

Given that knowledge implies belief, we can straightforwardly derive Williamson's margin for error principle **ME** from **SAF** and **B**.

Suppose that in α one knows that C obtains and $|v(\alpha)-v(\beta)| < c$ (where c is some small positive real number). Since knowledge implies belief, then in α one believes that C obtains. By principle **B**, then for some case β^*, $v(\beta^*) = v(\beta)$ and in β^* one believes that C obtains. Since β^* is close to α and by hypothesis one knows that C obtains in α, then by the safety condition **SAF**, in β^* one does not falsely believe that C obtains. Therefore, (by *modus ponendo tollens*), C obtains in β^*. Since, $v(\beta^*) = v(\beta)$, then C obtains in β. By a step of conditional proof, it follows that if in α one knows that C obtains and $|v(\alpha)-v(\beta)| < c$ then C obtains in β, which is effectively equivalent to the margin for error principle **ME**. [15]

---

[14] Contrast this with the stronger (and wholly implausible) principle which says that in all close cases one retains the same belief. Such a principle can feature as the induction step in a sorites argument.

[15] One can derive a different sort of margin for error principle from robustness plus the following principle:

(**F**) For all cases α and non-negative real numbers u, if $|u-v(\alpha)| < c$ and C obtains in α then, for some case β close to α, $v(\beta) = u$ and in β C obtains (where c is some small positive real number).

Informally, we can gloss of **F** as follows: if condition C obtains then it could easily still have been the case that C obtains if the parameter [upon which the presence or absence of condition C depends] had taken a slightly different value. If the tree is at most fifty feet high, then it could easily still have been at most fifty feet high if the tree had been differed in height by one inch, but not if it had been one hundred feet higher. Given the requirement of robustness encountered in the first chapter, which we write as follows:

(**ROB**) For all cases α and β, if β is close to α and in α one knows that C obtains, then in β it is not the case that: (one believes that C does not obtain and C obtains),

then the proof runs as follows: Suppose that in α one knows that C obtains and $|v(\alpha)-v(\beta)| < c$ (where c is some small positive real number). Since knowledge implies belief, then in α one believes that C obtains. By principle **F**, then for some case β^*, $v(\beta^*) = v(\beta)$ and in β^* C obtains. Since β^* is close to α and by hypothesis one knows that C obtains in α, then by the robustness condition **ROB**, in β^* it is not the case that both C obtains and one believes that it does not obtain. Therefore, (by *modus ponendo tollens*), one does not believe that C does not obtain in β^*. Since, $v(\beta^*) = v(\beta)$, then in β one does not believe that C does not obtain. By a step of conditional proof, it follows that if α one knows that C obtains and $|v(\alpha)-v(\beta)| < c$ then in β one does not believe that C does not obtain, which is effectively equivalent to the following margin for error principle (which I have indexed to a method of belief formation M):

(**MEB**) For all cases α and β, if β is close to α and in α one knows (via M) that a condition C obtains then in β one does not *believe* (via M) that C does not obtain.

There is structural parallel here between **ME** and **MEB**. The former principle shows that knowledge that p entails p not just in the actual case (i.e. factivity) but in nearby cases also. The latter principle shows that knowledge that p entails lack of belief that *not-p* not just in the actual case (as the condition **NBN** encountered in §1.5 demands) but in nearby cases also. Arguably, principle **F** stands or falls with principle **B**.

One immediate worry with this proof is its employment of the doxastic principle **B**. Is such a principle at all compelling? There seem to be at least two worries one might have with **B**. The first concerns its methodological pedigree, the second concerns its modal and epistemological status.

2.3 Principle B: methodology and status

The leitmotif of Williamson's *Knowledge and its Limits* is that knowledge comes first in the explanatory order, and indeed that knowledge is in some sense conceptually, and not just explanatorily prior (Williamson 2000, ch. 1-3). One consequence of this view is that knowledge is not to be explained in terms of any its allegedly component features, such as belief. It is rather that belief is to be explained in terms of knowledge. One then might wonder how well this methodology (which I confess I find very attractive) comports with Williamson's ambition to invoke the principle **B** in the derivation of the margin for error principle **ME** from the safety requirement on knowledge. Principle **B** is a purely doxastic principle, as such it doesn't seem like the sort of principle Williamson is entitled to employ in order to show that knowledge does not iterate or that luminosity fails. Only purely epistemic insights can tell us that **K** or **KK** fail if knowledge is explanatorily prior to belief

In reply to this worry, Williamson conceded that on a very 'pure', and perhaps canonical, reading of his methodology, this worry is well-founded but that one can nonetheless go some way to avoiding this worry by reconfiguring what we might mean by belief.[16] Very roughly, Williamson has suggested that one defines belief in epistemic terms as follows: a belief that p is an attitude to p, which for all one knows, is knowing p (*cf.* 2000, pp. 45-7). But might one know p while 'in a sense' treating p as if one did not know p (as in the nervous school-child example), and thus knowledge would not entail belief? Yes, but such cases are not paradigmatic cases of belief (i.e. cases of treating p as if one knows that p). Where belief is as defined, then knowledge indeed entails belief despite the fact that 'believing is not the highest common factor of knowing and *mere* believing, simply because it is not a factor of knowing at all (whether or not it is a necessary condition)' (*ibid.*, p. 47). Merely believing is just a kind of failed or 'botched' knowing, while knowing is the best kind of believing—without it being that knowledge has as a component *mere* belief. Williamson concedes that his epistemic

---

[16] Spoken reply at the symposium on his book *Knowledge and its Limits*.

account of belief is far from complete. But what if Williamson were able to avoid having to give such an epistemic account of belief in the first place? In §2.6, it is shown that we can adopt weaker margin for error principles which can do all the work of **ME** but without having to either rely on insights concerning belief.

What is the modal and epistemological status of **B**? The question is important since it will effectively settle the modal and epistemological status of **ME**. Presumably, the safety requirement, and the belief condition on knowledge, are both necessary and known *a priori*. If **B** shares this status, then **ME** will do likewise (given the plausible assumption that if the premises and the validity of an argument are known *a priori*, then the conclusion must likewise be known *a priori*, and the plausible assumption that the conclusion of a valid argument with necessarily true premises is itself necessarily true.). In a number of places, Williamson remarks that one can know *by reflection* that the methods via which we judge whether or not a condition C obtains are limited and less than perfectly discriminatory (e.g. Williamson 1994, pp. 219-20, 2000, p. 115). Such reflective knowledge is based on 'general considerations' about my capacity to see, taste, smell, and so on. Indeed, it is by reflection on such general considerations that we are supposed to grasp the truth of **B**. But knowledge gained via reflection is not thereby *a priori* knowledge, though some formulations of the *a priori* suggest this. Very roughly, S's knowledge that *p* is *a priori* if such knowledge is not based upon, or justified by, the character of (one or more) *particular* experiences had by S. That's consistent with the fact that in order for S to have *a priori* knowledge of a truth involving a concepts $c_1...c_n$, S must have had a sufficiently rich experience in order to grasp the content of each of these concepts (see Kitcher 2000, pp. 66-8). But it looks as if the grounds (or justification) for accepting **B** do indeed involve reference to the character of particular experiences.

I know that my eyesight is limited since after estimating by sight how many bricks there are in my garden, I was later to find out (by counting them) that my eyesight did not yield the right answer. I know that my ability to remember is less than perfect since I recall that there were two people in taxi even though at the time I knew damn fine that there were three. These particular experiences and countless like them go to form the very general considerations which lead one to have reflective but non-*a priori* knowledge of principle **B**. Since **B** is not known *a priori* then there are no good reasons to think that the margin for error

principle **ME** is known *a priori*. This is worrisome. If **ME** is not known *a priori*, it looks like in cannot feature as a *bona fide* axiom in our minimal theory.[17] But what of **B**'s modal status?

It doesn't seem at all plausible to say that **B** is an *a posteriori* but necessary truth. There are worlds in which my sensory perception produces beliefs which are never false—worlds where omniscient beings feed me the right beliefs. There are worlds in which I just happen to track the truth by pure accident: in such worlds I never form false beliefs, and there are worlds in which, for all p, I am inclined to believe that p in a case α but where I am not inclined to believe p in *any* close case β. Such reflections show that **B** should not be taken to be a necessary truth. Since the canonical ground for accepting **ME** is the derivation Williamson gives from **SAF** via **B**, then there are no grounds to think that Williamson's margin for error principles express necessary truths. So, it looks like principle **B** and principle **ME** do not have the right sort of modal and epistemological credentials to feature in a minimal theory of knowledge. Should we then simply weaken the demand that the platitudes of epistemic minimalism should record *a priori* conceptual truths? Arguably not.

One way to resolve this issue is to offer formulations of principle **B** and **ME** whereby we rigidify the name of the particular method employed. If 'method M' functions as rigid designator which picks out the same method in all α and β cases, then the following principles will surely be both necessary and knowable *a priori*:

(**B**) For all cases α and non-negative real numbers u, if $|u-v(\alpha)| < c$ and in α one believes that C obtains (via method M) then, for some case β close to α, $v(\beta) = u$ and in β one believes that C obtains (via method M).

(**ME**) For all cases α and β, if β is close to α and in α one knows (via method M) that a condition C obtains then C obtains in β.

If that's right, then this would seem to have consequences with respect to whether one can employ the margin for error principle **ME** in support of the realist thesis that there are undetectable truths. Consider the following argument:

(1) It is known (via M) that: C obtains in α and C does not obtain in β

(where α and β are close). Given the distributivity of knowledge over conjunctions this entails:

---

[17] Wright (1999, p. 226, fn.27) assumes that the relevant type of platitude exploited by truth-minimalism should preclude the use of non-*a priori* platitudes. Arguably, the same ought to hold for epistemic minimalism.

(2)   It is known (via M) that C obtains in α and it is known (via M) that C does not obtain in β

and so by &-E we infer:

(3)   It is known (via M) that C obtains in α
(4)   It is known (via M) that C does not obtain in β

given the truth condition **TC**, from 4 we can infer:

(5)   C does not obtain in β

and given the margin for error principle **ME**, from 3 we can infer

(6)   C obtains in β

Contradiction. So, reject 1 to infer:

(7)   It is not known (via M) that: C obtains in α and C does not obtain in β

Since, line 7 depends only on principles which are necessarily true, then by the rule of necessitation, and the modal equivalences, we can thus infer:

(8)   It is not *possible* to know (via M) that: C obtains in α and C does not obtain in β

What this shows is that there are unknowable truths. But is that not a distinctively realist thesis? This appears to be an unacceptable result. One would have thought that our minimal theory of knowledge is supposed to be neutral as to the question of whether there are verification-transcendent truths. This result might be a welcome consequence for Williamson who has in any case argued that Dummett's famous manifestation argument is undermined by his model of inexact knowledge (see Williamson 2000, pp. 110-113; *cf.* Williamson 1997b, pp. 908-9).

So we seem stuck with the disjunction: either our minimal theory of knowledge permits the derivation of a theorem which vindicates a realist conception of truth or our minimal theory of knowledge must dispense with margin for error principles since these principles vindicate a view which is simply too controversial. Since in §2.6, I also offer a margin for error principle (which is both *a priori* and conceptually necessary) from which it can be shown that there are unknowable truths via a proof similar to the one given above, then this dilemma is not confined to Williamson's model of inexact knowledge. This issue is not resolved until §3.8, where I argue that, while the above proof is sound, the conclusion provides absolutely no support for a realist conception of truth (or for a realist conception of

vagueness). Margins for error principles are available to both realists and non-realists alike, contrary to what might be thought.

So pending further argument, there is no reason to think that margin for error principles cannot feature as theorems (or indeed as axioms) in epistemic minimalism. However, more importantly for present purposes, there does seem to be good reason to reject not only the safety and robustness requirements on knowledge, but *any* form of local reliability condition. Let us see.

2.4 Lucky knowledge, local reliability, and Gettier cases

Not everybody shares the intuition that knowledge requires local reliability. Hetherington (1998) has recently argued that we should reject the credentials of modal epistemology outright and instead accept the traditional tripartite analysis of knowledge. On that basis, he argues that we indeed do have knowledge in Gettier cases, but a special type of knowledge, namely *lucky* knowledge—knowledge that we might easily not have had.[18]

There is certainly an anodyne sense in which much of our knowledge is lucky. It was only a matter of luck that I looked in the mirror and acquired the knowledge that newsprint was on my forehead. It was only by accident that Jones came to find out that he had heart disease (he bumped into a philanthropic specialist in the street who offered him a free consultation). Indeed, we have already seen that certain cases of lucky knowledge have implications with respect to some of the principles of modal epistemology. In particular, we saw in Chapter One, that married-Mary type cases show that Nozick's stability condition is unacceptable. But do cases of lucky knowledge undermine *all* forms local reliability conditions? And furthermore do all Gettier cases (where our belief is not based upon false evidence) amount to cases of lucky knowledge? Hetherington thinks so.

The essence of Hetherington's proposal is very simple: what happens in close counterfactual worlds is irrelevant as to whether a subject has knowledge. Only conditions which obtain in the actual world need to be considered. To take account of what would happen in nearby possible cases is to commit what Hetherington calls 'the epistemic counterfactuals fallacy' (*ibid.*, p. 456). Take the anti-luck Gettier case involving Henry in

---

[18] More specifically, he thinks that we do have knowledge in cases where our true belief is based on adequate evidence none of which is false. Feldman (1974) seems to have been the first to show that Gettier cases can arise even when one's (immediate) evidence is true.

38

barn-facade land. Henry has a justified true belief (not based upon false evidence) that he is seeing a barn and yet he lacks knowledge, so goes the usual story, since he could so easily have been mistaken—he could easily have formed the false belief that there was a barn in front of him since he could easily have looked at a barn-facade rather than a real barn.

For Hetherington, just because we lack knowledge in close counterfactual worlds does not undermine our actual knowledge. Being almost mistaken is not being mistaken, just like almost losing a race is not the same as losing the race. In both such cases we are lucky. Hetherington urges that we 'must must not link the actual and counterfactual too closely, or we shall confuse lacking knowledge with almost lacking it' (*ibid.*, , p. 459). So, any intuition that Henry lacks knowledge is ill-founded : it confuses nearly lacking knowledge with lacking knowledge. Hetherington says of Henry:

> His only undeniable epistemic failing is a counterfactual one, not an actual one. Admittedly, there is some temptation to interpret him as actually lacking knowledge. But it is easy to misjudge the worth of that temptation, which is rather like the misguided temptation to deny that an action is generous when our ground for that denial is that the person who performed the action is rarely generous in other situations, even in extremely similar ones, even in situations which he and/or we would regard as being qualitatively identical to the one where the action occurred. That denial would be mistaken, because even if the *person* is not generous, this does not entail that the particular *action* is not generous. Although fake barns do pose an epistemic threat to Henry (perhaps to his general trustworthiness in this area), this does not entail that they deprive him of knowledge right here, right now (1998, p. 458).

Bach (1985) has drawn an important distinction between a belief being justified and a person being justified in forming a belief. This distinction parallels the action-person distinction given by Hetherington. There is indeed a sense in which Henry's belief is justified, but also a sense in which Henry is not justified in holding this (justified) belief since the epistemic risk of forming a false belief concerning barns in barn-facade land is very great indeed. (So, the primary use of 'is justified' is as predicate of beliefs (or the mechanism which produce those beliefs) not of persons situated in a particular environment.) Once we recognise this distinction, there is an added reason to say that Henry has lucky knowledge

If that is right, then Henry knows that he is seeing a barn (he has a justified true belief and none of his immediate evidence is false) but he might so easily not have known this. It is an easy possibility that he should have formed the false belief that he is seeing a barn. This scenario is a straightforward counterexample to both the safety condition **SAF** and the sensitivity condition **SEN**. Lucky knowledge is incompatible with two of the four local

reliability conditions we considered in Chapter One. Since we have already rejected stability as a condition on knowledge (because of married-Mary cases), only the robustness condition remains. Does Hetherington reject this condition also? Indeed he does.

Hetherington reconsiders Harman's case of the political assassination and concludes that the subject Jill indeed has knowledge that the dictator has been assassinated even though she might easily have lacked this knowledge. (Presumably the news reports are generally reliable otherwise Jill's belief would not be well-supported by evidence.) Though it is easily possible for Jill to form a false belief (the false belief that the dictator has not been assassinated), this modal fact should not be taken to undermine the Jill's actual knowledge. Jill has lucky knowledge that the dictator has been assassinated. If that is right then this is a straightforward counterexample to both robustness and stability.[19] Jill has knowledge that the dictator is dead in the actual case, even though in nearby worlds where the dictator is assassinated, Jill not only fails to believe that he is dead she also believes that he is not dead. Are these observations at all cogent?

Hetherington is right to say that we have lucky knowledge not only in the two cases just presented but in all anti-luck Gettier cases (as I define them). But alas, his analysis does not extend to the standard types of Gettier examples. To substantiate these claims, firstly reflect on a further example. Take the case of Chisholm's dog cleverly disguised to look like a sheep (where we modify this case so as to exclude the worry over whether false evidence really is evidence). Every evening the farmer is concerned that some sheep may have got into his best corn field, so every evening he goes out to check. On one night, quite by a whim, his wife offers to go instead. She comes back and reports that there is a sheep in the field. The farmer thus has the following (true) immediate evidence 'My wife, who is generally reliable, tells me that there is a sheep in the field'.[20] On that basis, he forms the warranted belief that there is a sheep in his field. Indeed his belief his true: there is a sheep in the field but one which is lying

<hr/>

[19] Note that without the robustness condition one loses the direct route to the condition **NBN** given in §1.5. Thus, **NBN** must be retained on intuitive grounds alone.

[20] One might worry that a lemma in the farmer's chain of justification is the false belief 'My wife is telling me the truth'. But Hetherington does not demand that there is absolutely no false evidence in the chain of justification for the farmer's belief that there is a sheep in the field. So, there is no demand that the farmer's belief that his wife is telling the truth be a true belief nor a demand that the farmer's wife's statement 'There is a sheep in the cornfield' itself be based on true evidence (see p.460). The flip side of the demand that the evidence a subject does possess must not consist of *any* false beliefs is the demand that the evidence a subject does not possess must not consist of any true statements. Hetherington also rejects this demand since it would, for example, deprive Henry of lucky knowledge since there is an undercover fact (a true defeater of the form 'There are many fake barns here') which were Henry to become aware of this fact would defeat the original justification for his belief. On such accounts of justification, warrant entails truth (as Pollock acknowledges), but Hetherington wants to retain a fallible notion of justification in his tripartite analysis.

down out of sight in the far corner. What his wife took to be a sheep is in fact a dog which the farmer's son has mischievously disguised to look like a sheep.

Hetherington claims that in such cases the farmer indeed knows that there is a sheep in the field, but he is lucky to know this since were he to have gone himself he would have acquired a belief based on false evidence—he would have encountered the fake sheep in nearby cases and formed a belief on the wrong sort of basis, and so in close counterfactual cases he would not know. That is, in the counterfactual case, the farmer has the wrong basis for his belief that there is a sheep in the corn field since he infers this (true) belief from the false belief 'I am looking at a sheep'. Again, Hetherington assumes that one commits the epistemic counterfactuals fallacy in denying knowledge to the farmer.

Even so, Hetherington recognises that cases like this one work slightly differently to the Henry and Jill cases. In those cases, the standard allegation is that Henry and Jill lack knowledge because they form false beliefs in nearby cases. In the sheep case, in nearby cases the farmer does not form false beliefs, it is rather that he is alleged to lack knowledge because in nearby cases his justification is corrupt because it depends upon false evidence. That's a different sort of counterfactual failing. In giving his own diagnosis as to why it's wrong to say that a subject lacks knowledge in this case, Hetherington does not reject the principle that a properly justified true belief which is accidentally true cannot constitute knowledge. Rather, he rejects the principle that a properly justified true belief whose proper justification is only accidental cannot constitute knowledge (where proper justification means justification which does not depend on any immediate false evidence). So it's important to distinguish two types of anti-luck Gettier case: the type where one is alleged to lack knowledge because one could easily have been in error in close cases, and the type of case where one is alleged to lack knowledge because one could easily have used the wrong sort of evidence. In both types of anti-luck case, one lacks knowledge in nearby cases, though for different reasons in each case.[21] But surely we are inclined to say that the farmer does not have knowledge (lucky or otherwise) that there is sheep in the corn-field. Before substantiating this response, let me outline the case that Hetherington himself employs.

Hetherington does not use Chisholm's sheep case as an example of accidental justification but rather the so-called 'existential generalisation case' given by Feldman (1974). In this

---

[21] Reed (2000) highlights a range of Gettier cases involving accidental justification and, unlike Hetherington, takes these cases to be much more harmful than cases of accidental truth. Reed does not recognise the possibility of lucky knowledge nor does he recognise that the two types of anti-luck cases can overlap when one has justification which is both accidental and factive.

case, Smith has the true belief that someone in the office owns a Ford. His belief is true because Havit, his office-mate, owns one. Smith's belief is justified because Nogot has claimed to own a Ford and Smith has seen Nogot's certificate of ownership, and indeed Nogot has always been trustworthy in the past. So, Smith's evidence is 'There is someone in the office who has always been honest with me and claims to own a Ford and indeed they have shown me a certificate to that effect'. This is indeed true evidence. Intuitively, Smith does *not* have knowledge that someone in the office owns a Ford. But why? Hetherington suggests that the reason why we have this intuition is that we are calling upon something like the following reasoning:

> If Smith had reasoned slightly differently—in particular, if he had inferred the same conclusion from [the false statement] 'Nogot owns a Ford', instead of from the existential generalisation he did use—he would unwittingly have been misled in his reasoning. He would have been using false evidence, decreasing his chances of deriving a true conclusion. And Smith *might*, so easily, have reasoned in that way. So he *would* have reached the same true conclusion ('someone in the office owns a Ford') without realising that he was being misled along the way. Hence, his reaching that conclusion now, also without thinking that he is being misled, does not make his belief knowledge, even though it is true and he is not actually being misled [...] The counterfactual lack of knowledge (if he had used false evidence in his reasoning) implies the actual lack of it (when he actually uses no false evidence) (1998, pp. 455-6).

Hetherington thinks that it is irrelevant whether Smith got his evidence in the wrong way in close counterfactual cases. Since he did not get his justification via any false lemma as things actually stand then Smith can be said to have lucky knowledge.

But surely just as we are highly disinclined to say that the farmer has knowledge (lucky or otherwise) that there is a sheep in the corn-field, we are also highly disinclined to say that Smith has knowledge (lucky or otherwise) that someone in the office owns a Ford. Moreover, one can offer a diagnosis as to why the farmer and Smith lack knowledge without appealing to the principle that a justified true belief which is accidentally justified (in the right sort of way) cannot constitute knowledge. But what is the relevant difference between the anti-luck cases of Henry and Jill and the more standard Gettier cases involving the farmer and Smith?

Hetherington omits to discuss or even mention that in the Henry and Jill Gettier cases the subject's belief is causally connected to the facts. Is that significant? The fact that there is a barn causes Henry to believe that there is a barn (given his disposition to form beliefs about sundry salient objects in his immediate environment). Likewise in the assassination case, the news report (together with the subject's disposition to believe what is heard on the news) causes the subject to believe that the dictator has been assassinated. In the sheep-case, in contrast, there is no causal connection between the fact that there is a sheep in the field and the farmer's belief. The farmer's wife does not notice the real sheep and her testimony is based on seeing the fake sheep. Likewise in the case of Smith. The fact that someone in the office owns a Ford bears no causal relation to Smith's belief. This is crucial. When the causal connection is absent then we are far less inclined (if indeed inclined at all) to say that either the farmer or Smith has knowledge (lucky or otherwise). At the very least, it looks like not all cases of true belief which is well-supported by evidence, none of which is false, can count as knowledge.[22]

Hetherington allows that a true belief which is luckily justified in the right way can count as knowledge even when there is no relevant connection between the fact and one's belief or the reasons for one's belief. But surely, it's the most natural and obvious diagnosis to say that when there is no causal condition present, when one's belief (or the reasons for one's belief) bear no causal relationship to the facts, then knowledge is ruled out. The possibility of lucky knowledge, and the attendant rejection of all forms of the local reliability condition, ought not lead us back to the traditional tripartite analysis of knowledge, whereby a justified true belief just is knowledge (where one's justification is both fallible and not based upon false evidence). Rather, it ought to lead us back to an account of knowledge which sanctions the relevant connection condition we encountered in the first chapter:

(RC)   Necessarily, if S knows that *p* then S's belief that *p*, or S's reasons or justification for their belief that *p*, are relevantly connected to the fact that *p*

---

[22] One of the key theses of Hetherington's paper is that fallibilism is compatible with a JTB-analysis of knowledge once we reject modal epistemology. Even if we can rescue a JTB-analysis in the barn-facade and assassination cases by requiring that justification have a causal condition built in this nonetheless fails to sanction fallibilism. If one's belief has to be caused by the facts in order to be justified then there is no scope for one's belief to be justified but false. Lucky knowledge, properly conceived, cannot rescue the conjunction of fallibilism and a tripartite analysis of knowledge.

Here the relevant connection is not to be read in terms of any local reliability condition, but is rather to represent a causal connection of the appropriate sort.

Once we allow for lucky justification, it's a small step to allow for a lucky causal connection (of the right sort) between the fact that p and the subject's belief that p (or the reasons for their belief). But notice what is being claimed here. I am not trying to explain why Henry and Jill have (lucky) knowledge by adverting to the presence of a causal condition. I'm bracketing all consideration of what constitutes the jointly sufficient conditions for knowledge. Indeed, I'm merely appealing to the intuition that Jill and Henry do have (lucky) knowledge. Rather, I'm employing the condition **RC**, when read as specifying a causal connection, to explain why the farmer and Smith lack knowledge. In contrast, Goldman (1967) intended the causal condition to feature as a component in a jointly sufficient condition for knowledge.[23] However, our minimal theory is only concerned with necessary conditions which enable us to predict the presence of ignorance not the presence of knowledge.

Of course Goldman (1976) put forward the barn-facade case as just the sort of case that his earlier causal theory was unable to handle since he thought it intuitively obvious that Henry lacks knowledge. Since the causal theory is unable to diagnose just why this is so, Goldman concluded that we must reject it. Even though his diagnosis has been very influential, was he right to jettison his earlier proposal so readily? Might the force of such examples have been overstated? Arguably so, given the general tendency of modal epistemology to overlook any form of lucky knowledge. Once we make room for lucky knowledge in conceptual space, then there is very good reason to re-think just why the causal theory of knowledge was discarded.

To defend a causal condition on knowledge would require more space than is available here and would indeed take us beyond the strict remit of epistemic minimalism. However, since I am suggesting that the relevant connection principle **RC** (when read causally) be re-admitted into our minimal theory of knowledge at the expense of all forms of local reliability condition, then some discussion of the plausibility of this move is required.

Here is a list of immediate worries that one might have with respect to admitting a causal version of the principle **RC** into a theory of knowledge (minimal or otherwise) which admits the possibility of lucky knowledge:

---

[23] In fact, Goldman (1967) suggested that the causal condition be a fourth condition in addition to truth, justification, and belief. (Strictly speaking one can drop the truth-condition, when one demands that one's belief be hooked up to the facts.)

(a) Even if the causal version of **RC** is a minimal axiom, the supposition that one can know even when the causal connection could easily have been absent is incompatible with the natural idea that knowledge is a cognitive achievement which typically demands epistemic praise.

(b) If the causal version of **RC** is a minimal axiom then it follows that epistemic minimalism is committed to epistemological externalism since from the inside one cannot tell the difference between cases where there is an appropriate causal connection present and cases where there is not.

(c) It is too drastic to reject all forms of the local reliability condition, for then one must reject the credentials of modal epistemology altogether.

(d) If the causal version of **RC** is a minimal axiom then we shall have no way to account for our knowledge of facts which have no overt causal powers and properties, such as mathematical knowledge, logical knowledge, and perhaps knowledge of universal generalisations.

(e) Any plausible account of causation will have to support counterfactual conditionals. In so doing, such an account will inevitably entail the sensitivity condition on knowledge. Thus, the causal theory rules will be incompatible with certain cases of lucky knowledge contrary to the advertised features of the theory.

(f) It is not possible to specify when a causal connection is appropriate (or 'relevant') and when it is not. Without such a specification, the principle **RC** is useless.

Let me briefly take these worries in turn, beginning with (a).

Knowledge is typically seen to be a condition which is highly prized, and indeed to gain knowledge is generally seen to be an achievement. One might even use the metaphor that belief is like shooting at a target with knowledge as the highest form of belief represented by the centre-circle. To hit the centre of the target by accident (a slight breeze moved one's arrow) doesn't feel like the sort of thing that should be praised. But everyone can agree with that. Cases of lucky knowledge are not completely accidental. For one thing, in order to gain the right belief in the barn-facade case, one has to have exercised one's perceptual apparatus and method of belief formation properly. The tendency to withhold any form of approbation for lucky knowledge looks like it must then stem from the feeling that one could just as easily have been doing one's epistemic duty in judging p, and indeed have satisfied the global reliability requirement **GR**, and yet have formed a false belief (by looking at a fake barn, say). In such cases, one might claim that while one remains deontologically justified in believing p (one is epistemically blameless) there is no additional praise for getting the matter luckily right—in the sense of forming a belief which is accidentally connected to the facts in the appropriate fashion. It is this additional praise, so goes the challenge, that is required for knowledge. But is that right?

45

One might argue that accidental success does receive praise—some athletes are genetically predisposed to do better than others who train equally as hard and yet we praise them far more because they win. A better response is to question the presupposition that knowledge that *p* necessarily requires more praise than merely doing one's epistemic duty in forming the (true) belief that *p*. Some of our knowledge is just so lucky (a benevolent demon gives me x-ray specs and I notice that there is a bomb in the bag, or the fact that *p* just happens to cause to me to believe that *p*) that there is no real sense in which it is praiseworthy.[24]

Is it useful, then, to distinguish between what we may term low-grade knowledge and high-grade knowledge? In cases of low-grade knowledge, one's knowledge is lucky in the sense that one gets things right but in so doing one has subjected oneself to great epistemic risk in the sense that one could easily have got things wrong (in fake-world, for example). High-grade knowledge is where one get things right but there is no attendant risk when forming one's belief (in real-world, for example).[25] While this distinction is worthwhile, the difference between real-world and fake-world is not discernible by the subject. The distinction between person-justification and belief-justification drawn by Bach (1985) is relevant here (see previous section). In cases of lucky of knowledge a *person* can be said to unjustified in forming a belief due to the attendant epistemic risk. That's why it's a low-grade form of knowledge. Even so, to have knowledge in fake-world is also in a sense just as praiseworthy as having knowledge in real-world—one has done one's epistemic duty in both cases, and one has managed to form a true belief. The fact that one's belief was much less risky in real-world is a feature of the knowledge-friendliness of the environment, it is not due to any special cognitive achievement on behalf of the subject. Hence, the subject, ought to receive no extra praise in such a case. I tentatively conclude that objection (a) is groundless.

The above remarks lead us to the worry stated in (b). Roughly, the worry here is that satisfaction of the causal condition will not typically be something which is reflectively accessible to the knowing subject. Hence, whether one knows will not itself be something which is reflectively accessible to the knowing subject. Underpinning this worry is the idea a subject cannot discriminate between any of the three following cases: when they have low-grade, lucky knowledge, when they have high-grade knowledge, and when they lack

---

[24] Indeed, much of our knowledge is what we may call *easy* knowledge—knowledge that is not easy to lack (see Zagzebski 1999, p. 94-5). When conditions are good, and I am looking at a red object, and I wonder whether there is a red object in my line of vision, then it is not easy to lack knowledge that there is a red object in such a case.

[25] The title of Hetherington's forthcoming book *Good Knowledge, Bad Knowledge*, Oxford: Oxford University Press (which I have not yet had the chance to read) suggests a similar distinction.

knowledge altogether.[26] It thus looks like, in a great many cases, a subject will not be in a position to know (by introspection or reflection) that they know. Is there a response? Notice that what is driving this worry is, in part, the thesis that knowledge requires discrimination. Typically, that thesis is cashed out in reliabilist terms using some form of a local reliability condition on knowledge (Goldman 1976 is the paradigm example). But once such local reliability conditions on knowledge have been rejected then, for example, there is no requirement that Henry be able to discriminate barn-world from barn-facade world in order to have second-order knowledge. At the very least, this opens up the possibility that Henry is in a position to know that he knows that he is seeing a barn. So, principle **PK** and principle **K** are not necessarily threatened by the analysis in hand. What does seem to be threatened is the internalist requirement that in order to have second-order knowledge one must be able to reflectively access that one has first-order knowledge. Perhaps even this requirement can be accommodated. How so?

Supposing a subject forms the second-order belief that their first-order belief that they are seeing a barn constitutes knowledge. What justification can they cite for their second-order belief? They could cite the following reasons: that they have been generally been reliable in the past, and indeed that their most of their past beliefs have always been relevantly connected to the facts. If the reasons for these beliefs bear the appropriate relationships to the facts (the epistemic facts in this case) then there is a route for the subject to be able to reflectively access the fact that they know that they know. If that is right, then it is not obvious that a minimal theory containing the causal axiom **RC** entails an overtly externalist epistemology.

With respect to (c), even if one rejects all forms of the local reliability condition there are at least three ways in which one might retain some form of modal epistemology. Firstly, even Hetherington recognises that one might retain a modal element in one's theory of knowledge. He says that his view

does not leave counterfactuals with no epistemological contributions to make. To mention but two examples: perhaps they help to determine whether an epistemic subject really believes that *p*, or how well he is using his evidence (1998, p. 466).

---

[26] I should add that given Hetherington's diagnosis of lucky knowledge cases, and his attendant defence of the traditional tripartite analysis of knowledge, he thereby wishes to defend both fallibilism and internalism. Thus, the present worry does not afflict his own model of lucky knowledge. For more on Hetherington's model of fallible (what he calls 'failable') knowledge, see his (1999).

Hetherington does not elaborate, but his remarks are suggestive. Rather than speak of *belief* as being the bearer of the properties of sensitivity, stability, safety, and robustness (and their complementary properties), one ought to speak of *knowledge* as being the primary bearer of these properties. Hence, counterfactual conditions can tell us, amongst other things, just how lucky or non-accidental our knowledge might be. These conditions can be used to distinguish various ways in which our knowledge may be low-grade or high-grade. Again, this helps explain why it is so easy to confuse nearly lacking with lacking knowledge: we simply confused very low-grade knowledge with ignorance.

But while Hetherington can retain a modal element in his epistemology, he is nonetheless adamant that counterfactual conditions are irrelevant with respect to whether a subject has or lacks knowledge. In the next section, I will argue that this extreme view is mistaken, that there are minimal margin for error principles which are modal in nature and which even Hetherington is beholden to accept. Furthermore, even if one does accept the possibility of lucky knowledge this does not entail that a local reliability condition has no role to play in an analysis of knowledge. This connects with the problem (d)—that a causal condition rules out the possibility of knowing facts which are non-causal in nature.

Though the analysis of knowledge I have been defending dispenses with safety and robustness as self-standing conditions of knowledge, there is no obstacle to retaining them as part of a larger disjunctive condition, which we might give as follows:

(D) $(\forall \alpha)$ $(\forall \beta)$ if β is close to α and in α S knows that C obtains then *either*

   (i) (in β, it is not the case that : S believes that C obtains and C does not obtain) and

      (in β, it is not the case that: S believes that C does not obtain and C does obtain), *or*

   (ii) S's belief that C obtains (or their reasons for this belief) is caused (in α) by the

      fact that C obtains (in the right sort of way).

Call **D** the disjunctive condition on knowledge. This principle would allow the causal condition in the second disjunct of the consequent to fail in certain cases but allow that knowledge is retained. That would have the advantage of ensuring that the causal condition does not threaten knowledge of facts which, on certain conceptions at least, have no obvious causal impact such as mathematical facts, necessary truths, and universal generalisations.[27] Notice also that **D** correctly predicts the failure of knowledge both in the (modified) case of

Chisholm's sheep and the existential generalisation Smith-Nogot case. In these cases, both disjuncts of **D**'s consequent fail to hold and thus the subject lacks knowledge. In contrast, **D** does not fail in the Henry and Jill cases, for in such case only the first disjunct of the consequent of **D** fails. This at least suggests that **D** may be along the right lines.

With respect to (e) it does indeed appear that in order to specify the causal condition properly one may well have to rely on some form of local reliability condition. Armstrong (1973, pp. 162-75; pp. 178-83) proposed a model of (simple perceptual) knowledge whereby, roughly, a belief that p counts as knowledge if it is both true and bears a law-like connection to the fact that p. For Armstrong, such law-like connections will support the following counterfactual claim: 'if p had not been the case then it would not have been the case that A believed that p' (*ibid.*, p. 166).[28] A naive causal theory, in contrast, simply says if one knows that p then one's belief that p must be caused by the fact that p (in the right sort of way). Yet even on the naive model it's not obvious how the sensitivity condition can be avoided. If that's so, then it's hard to find room for the phenomenon of lucky knowledge within a causal theory of knowledge. Indeed Lewis (1973, p. 557) takes the following remarks to be 'platitudes':

> we do know that causation has something or other to do with counterfactuals. We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it. Had it been absent, its effects—some of them at least, and usually all—would have been absent as well.

But is that really a platitude? Suppose in fake-world, there are rare but real alpine flowers that bloom for only a few seconds then wilt. Immediately after they wilt, a projector projects a hologram of a real flower in order to extend the (appearance of) the flowering season of these rare and beautiful flowers. Jimmy wanders into the countryside of fake-world. He luckily sees a real alpine flower and on that basis forms a belief about what he sees. In this case, the fact that there is a real flower causes him to believe that there is a real flower. However, in the nearest world in which there is no real flower but a projected one he continues to form a belief that he seeing a real flower. Hence, Lewis's 'difference-principle' fails in this case,

---

[27] Hetherington could allow a disjunctive condition also by replacing the causal condition with a justification condition.

[28] Armstrong's model is not strictly a causal model since while he holds that the law-like connection between the fact that p and one's belief that p will 'frequently' reflect a causal connection between the fact and the belief, it may not do so in all cases. Causal-reliability accounts (e.g. Goldman 1976) do not posit a connection between the fact that p and the belief that p they simply require that one's belief be produced/caused by a reliable process/method.

even though his actual belief is caused by a real flower. Such an example, and many like it, highlight how difficult it is to state the conditions under which a belief has been caused by the facts (in the right sort of way). They also show that there is no direct route from a causal condition on knowledge to the sensitivity condition **SEN**. Perhaps we merely have to rely on a primitive account of when a belief is caused by the facts. If such a move is allowable then we may conclude that lucky knowledge is compatible with a causal condition on knowledge. What then about problem (f)?

Perhaps the most pressing worry with any causal analysis of knowledge is the proviso that one's belief must be caused in the right sort of way—the right portions of reality must be causing one's belief. If one cannot specify this proviso properly, then the challenge runs that there are obvious counterexamples to hand. Take the sheep case. Suppose the reason that the farmer's son disguised a dog as a sheep and put it in the field was in order to give the real sheep some company. The presence of the real sheep causes the disguised sheep to be there which causes the farmer's wife to say that there is a sheep in the field. But this looks like the wrong sort of causal connection between the facts and the farmer's belief. One response to such difficulties is to demand that the causal chain does not involve or 'pass through' any *misleading defeaters*—facts which, were one to become aware of them, would cause you to withhold your belief (or would undermine any evidence you have for your belief). That would deal with the counterexample just sketched, since the causal chain involves a fact (the fact that there is a dog cleverly disguised as a sheep in the field) which were you to become aware of this fact would defeat the grounds for your belief. However as Harman (1973, p. 152) has argued, misleading defeaters are ubiquitous, so it's not clear how useful this specification might be. In virtue of the cottage-industry generated by Gettier cases, a certain pessimism is in order.[29] Perhaps the best response to this problem, is to reject the demand that one give a specific characterisation of what counts as the appropriate sort of causal connection and rest content with our intuitions in individual cases. This would mean that we keep the causal variant of relevant connection principle **RC** as one of Williams's 'highly formal remarks' which can only be successfully applied by having a sensitivity to the various contextual factors which are relevant case by case. Such a predicament in any case chimes with the manner in which we accepted the global reliability condition **GR** in Chapter One.

While the above responses are programmatic and incomplete, I hope I have done enough to show that a causal condition on knowledge is not without merit, that such a condition is

compatible with lucky knowledge, and that a causal theory of knowledge is not quite the dead duck that many have taken it to be. What then of the credentials of modal epistemology—is Hetherington right to think that what happens in nearby possible worlds is absolutely irrelevant as to whether a subject has or lacks knowledge?

2.6 Minimal margins for error, luminosity, and knowing that one knows

Consider the following margin for error principle (which for reasons I hope will become apparent in due course I call *minimal* margin for error principles):

> (**MME**) For all cases α and β, if β is close to α and in α one knows (via M) that a condition C obtains then in β one does not know (via M) that C does not obtain.

(where M is a rigid name for the method the subject employs in order to acquire knowledge).[30] Take the case of the tree and the condition of being less than ten feet tall. The principle **MME** says that knowledge that the tree is less than ten feet tall precludes knowledge that the tree is not less than ten feet tall in close cases (actual or counterfactual).

One key feature of **MME** is that it is compatible with cases of lucky knowledge. In the case of Henry in barn-facade land, Henry's method of acquiring knowledge is such that he cannot discriminate real barns from fake barns. This does not however preclude him from having lucky knowledge that he is seeing a barn in knowledge-unfriendly environments such as barn-facade land. In a case α, it appears to Henry that he is looking a barn; in a close case β, it also appears to Henry that he is looking at a barn. These two cases are doxically and perceptually identical, even though let us suppose that in the α case Henry forms a true belief that there is barn in his line of vision, while in the close β case, Henry forms a false belief that there is a barn in his line of vision. The margin for error principle **ME** predicts that Henry lacks knowledge: his true belief in the α case cannot be knowledge as in the β case it is false that there is a barn in his line of vision. (One could equally well predict that Henry is ignorant in the α case by using the safety requirement as we have done above.) However, the margin for error principle **MME** does not entail that Henry lacks knowledge in the α case, since Henry would have to know that he is not seeing a barn in the doxically identical case β.

---

[29] See Shope (1983, Ch.5) for a good introduction to some of the problems faced by a causal analysis of knowledge (and descendants).

Clearly, Henry lacks knowledge that he is *not* seeing a barn in the β case. But why exactly? The simplest explanation is to say that Henry does not form the belief that he is not seeing a barn in the close case β. Given the belief condition **BC**, Henry thus does not know he is not seeing a barn. Thus, Henry's (lucky) knowledge in the α case is not ruled out.

The principle **MME** is a very weak modal condition, but it can serve to ground a model of inexact knowledge. Hetherington is simply wrong to reject the idea that what happens in nearby cases is irrelevant with respect to whether a subject knows. Indeed, given our discussion of lucky knowledge, and the good reasons we found to reject all forms of the local reliability condition, it may well turn out that **MME** represents one of the few principles which can feature in a plausible modal epistemology.

MME is weaker than **ME**. Why? Given the factivity of knowledge, it's trivially true that if a condition C obtains then one does not know that C does not obtain, as given by the following schema:

(M) $(\forall \alpha)\, C_\alpha \to \sim K[\sim C]_\alpha$

Given transitivity plus **M** one can straightforwardly derive **MME** from **ME**. However, the converse entailment does not hold: one cannot derive **ME** from **MME**.[31]

In many cases, where a weaker principle will do the work of a stronger principle we have reason to invoke the former rather than the latter principle. To overturn such a preference, the burden of proof is on the devotee of the stronger principle to show that the only or indeed the best motivation for the weaker principle comes via the stronger principle which entails it. But equally to avoid any danger of stalemate, there also seems to be a burden of proof on the devotee of the weaker principle to show both that the weaker principle has independent motivation *and* that the stronger principle is less motivated than the weaker (or indeed perhaps lacks motivation altogether). We have already advanced considerations against **ME**, by rejecting the soundness of Williamson's proof from safety (via **B**) to **ME**. Hence, there are no grounds to accept **MME** on the basis of accepting **ME**. Indeed, it looks like we can provide an independent motivation for **MME**.

---

[30] Thus, **MME** is both necessary and knowable *a priori*.

[31] Given the belief condition on knowledge, one can also derive **MME** from the margin for error principle **MEB** given as follows: For all cases α and β, if β is close to α and in α one knows (via M) that a condition C obtains then in β one does not *believe* (via M) that C does not obtain. Since knowledge entails belief (such that lack of belief entails lack of knowledge) then given the transitivity of the conditional, we can straightforwardly show that **MEB** entails the minimal margin for error principle **MME**.

Principles like **MME** I call *minimal* margin for error principles because it looks as if we can motivate them on grounds which are theoretically fairly lightweight. Even so, it looks like there are two minimal motivations for **MME** on offer—the *descriptive* and the *explanatory*, the former being theoretically more lightweight than the latter.

The descriptive route to **MME** begins with insights drawn from exemplars of inexact knowledge. Consider a stadium gradually filling up with people. By looking, there are no close instants in time such that I know at one instant that the stadium contains less than 39,000 people and I know at the next instant that the stadium does not contain less than 39,000 people. However this result merely shows that **ME** holds for a particular condition C. We need **ME** to hold (schematically) for all relevant conditions C (which includes epistemic conditions with any number of iterations of knowledge). To achieve that we must extrapolate by employing a meta-principle which Williamson gives as follows:

> Where knowledge is inexact, some margin for error principle holds [...] inexact knowledge is a widespread an easily recognisable phenomena, whose underlying nature turns out to be characterised by the holding of margin for error principles (1994, p. 227).

It's natural to employ this meta-principle to generalise over particular cases and offer the generic insight that where we have some condition C, whose application depends on the variation in some graded or continuous parameter, you will find that no small variation in this parameter permits knowledge that C obtains and knowledge that C does not obtain across such a variation. Arguably, such an insight can be taken to conceptualise a basic epistemic principle, namely **MME**, which is itself explanatorily prior to further insights concerning the nature of knowledge. Such thoughts comport extremely well with the *knowledge-first* methodology sponsored by Williamson.

That said, such a *descriptive* motivation for minimal margin for error principles is likely to seem somewhat theoretically unsatisfactory. Just *why* do such minimal margin for error principles govern (inexact) knowledge? Williamson himself gives part of the answer when he says that the '[m]ain idea behind the argument against luminosity is that our powers of discrimination are limited' (2000, p. 13). In and of itself, such a remark does not advert to 'limited discrimination in the belief-forming process' (*ibid.*, p. 127), nor should it. We can simply claim that our knowledge that C obtains cannot sit side by side with knowledge that C does not obtain (in close cases) *because* the methods by which we acquire *knowledge* that C obtains are less than perfectly discriminatory. A model of (inexact) knowledge of this sort is indeed compatible with the thesis that knowledge may not after all entail belief. It would also

seem to be compatible with the 'pure' reading of Williamson's methodology canvassed above, such that doxastic insights cannot inform claims concerning knowledge.

Given that a case has been made for the minimal credentials of **MME**, can we employ this margin for error principle to undermine both transparency and iterativity? This is the final and most important question which we shall ask in this chapter. **MME** is equivalent to the negative existential:

1 (1) $\sim(\exists\alpha)(\exists\beta) K[C]_\alpha \,\&\, K[\sim C]_\beta$ (where β is close to α).

(where '$K[C]_\alpha$' abbreviates 'In α one knows that condition C obtains', and where '$K[\sim C]_\beta$' abbreviates 'In β one knows that C does not obtain'). Plausibly, the claim stated in (1) is available to be known by the knowing subject. (As has just been mentioned, the knowing subjects reflects on the fact that his methods of acquiring knowledge are less than perfectly discriminatory.) Hence:

1 (2) $K[\sim(\exists\alpha)(\exists\beta) K[C]_\alpha \,\&\, K[\sim C]_\beta]$ (where β is close to α).

Now suppose for the sake of argument the following:

3 (3) $K[\sim C]_\beta$
4 (4) C_α (where β is close to α)

Also suppose that we have a K-introduction rule in the following form:

$$\frac{\Gamma \vdash C_\alpha}{\Gamma \vdash K[C]_\alpha}$$ (K-introduction)

This rule is valid in two scenarios: (i) where the condition C is a Cartesian condition and the knowing subject has properly considered whether C obtains, and (ii) where C is the epistemic condition of knowing p, and where the knowing subject has properly considered whether they know p, *and* where all the formulas in Γ are known.[32] This latter clause ensures that one knows that one knows that p only if one's knowledge that p does not depend on any unknown premises. The K-introduction rule stands or falls with the principle **K**.

Since by hypothesis, C schematises both Cartesian conditions and the epistemic condition of knowing that p, then we can apply the K-introduction to line 4 to infer:

---

[32] More specifically, any formula in Γ must be what may be termed KT4-fully modal. A wff of the form '$K\phi$' is KT4-fully modal, and if 'ϕ' and 'ψ' are fully modal in this sense then so are '$\phi \,\&\, \psi$' and '$\phi \lor \psi$'.

4 (5) $K[C]_\alpha$

and by &-I on lines 5 and 3, together with the rule of existential-introduction, we can then infer:

3, 4 (6) $(\exists\alpha)(\exists\beta) K[C]_\alpha \ \& \ K[\sim C]_\beta$ (where β is close to α)

Given factivity, lines 2 and 6 contradict, and so we can reject the assumption at line 4 to infer:

1, 3 (7) $\sim C_\alpha$

Given the plausible assumption that if a condition C is Cartesian then the complement of that condition is likewise Cartesian (together with the fact that one has considered whether C does not obtain) means that clause (i) of the K-introduction rule is satisfied. Furthermore, since line 7 rests upon formulas which are prefixed by 'K' (together with the fact that one has considered whether one know that C obtains) then clause (ii) of the K-introduction rule is satisfied. Either way then, we can infer:

1, 3 (8) $K[\sim C]_\alpha$

and by conditional proof, and two applications for universal instantiation, we infer the undesirable schema:

1 (9) $(\forall\alpha) (\forall\beta) K[\sim C]_\beta \rightarrow K[\sim C]_\alpha$ (where β is close to α)

which is sorites-generating. Take the condition of not feeling hot. If this condition is known to hold in one case then it holds in every case. Likewise, take the condition that the tree is less than ten feet tall. There are cases where one knows that this condition fails to obtain (after it grows to a height of twenty feet tall), yet one can use the unpalatable schema to derive the absurd result that one knows that this condition fails to obtain for all possible heights of the tree. Conclusion: reject the rule of K-introduction when the condition C is Cartesian or when it is the condition of knowing p. Given the deduction theorem, that's just to reject transparency and iterativity.[33] Thus, given the minimal margin for error principle **MME** it follows that internalism with respect to self-knowledge and internalism with respect

---

[33] This proof is based upon Wright's so-called paradox of higher-order vagueness, though with significant adjustments (Wright 1987, 1992). We re-use this proof in a different way in §3.9.

to knowledge in general are both untenable doctrines. From axioms which everybody should be prepared to endorse, it follows that we must reject the two forms of internalism in hand.

What have we achieved in this chapter? We have found good reason to reject all forms of the local reliability condition using arguments derived from Hetherington (1998). For that reason we found it impossible to incorporate Williamson's model of inexact knowledge into our minimal theory of knowledge. But it was also found that Hetherington is mistaken to extend the class of lucky knowledge cases to all forms of Gettier case. Instead of rehabilitating the traditional tripartite analysis of knowledge, our investigations led us to rehabilitate a causal version of the relevant connection condition encountered in Chapter One. Moreover, we also found good reason to reject Hetherington's actualist epistemology on the grounds that knowledge is governed by the minimal margin for error principle **MME**. Just like **ME**, this principle can be employed to undermine internalism with respect to knowledge and internalism with respect to self-knowledge. Since these principles are axioms of our minimal theory, then we were able to show that, in two key respects at least, we are all externalists now.

CHAPTER THREE
THE MINIMAL THEORY OF VAGUENESS[1]

Hitherto, there has been a consensus that a constitutive definition of vagueness is too much to ask for. The aim of this chapter is to challenge that consensus by identifying the constitution of vagueness from a perspective which is as neutral as possible on matters both logical and philosophical. In so doing, I shall lay the foundation for *a minimal theory of vagueness*. To that end, three related dimensions of vagueness are distinguished: vagueness *qua* sorites-susceptibility, vagueness *qua* borderline cases, and vagueness *qua* tolerance. Hitherto, the relationship between these dimensions has remained somewhat unclear. The minimal theory of vagueness is equipped to remove much of that unclarity. One overall merit of this theory is that it promises to ensure that the dialectic of the vagueness debate can at least begin at a mutually agreed point—this theory can at least ensure that we are all taking about the same phenomenon from the outset. Like the account of knowledge developed in the previous chapters, the axioms of this theory are intended to be as uncontroversial as possible while some of its theorems, as we shall find in §3.9, are decidedly controversial.

As a preliminary to such investigations, we will address what we can reasonably expect or demand from a minimal theory of vagueness. Can this theory solve the sorites paradox? Can it isolate the source of linguistic vagueness? Can this theory successfully rehabilitate the so-

---

[1] This chapter is based on a talk given at a workshop on vagueness held at the Institute of Philosophy, University of Bologna, 22nd-23rd November, 2001. Many thanks to the audience on that occasion for their helpful feedback and particularly to Andrea Sereni for his stimulating reply.

called *characteristic sentence approach* to defining vagueness? These are the sorts of questions addressed in §3.1. In §3.2, it is found that vagueness defined as sorites-susceptibility offers the least controversial characterisation of vagueness. However this characterisation proves to be too insubstantial for the promises of the minimal theory to be properly satisfied. On what is perhaps the most prevalent conception, vagueness is the phenomenon of borderline cases. From §3.3 through to §3.5, I assess whether it is plausible to give an uncontroversial definition by reference to such a phenomenon. A number of non-epistemic and epistemic accounts of what it is to be a borderline case are scrutinised. For the purpose of finding a neutral definition of vagueness, none of these proves entirely satisfactory (the particular bug-bear proves to be the possibility of terms which we can stipulate to give rise to borderline cases but which draw sharp and clearly identifiable divisions across their associated dimension of comparison). *Prima facie*, it is far more plausible to minimally define vagueness by reference to an epistemic notion of *tolerance*. Such a notion is intended to capture the thesis that vague terms draw no clear or known boundary across their range of signification and contrasts sharply with the (semantic) notion of tolerance given by Wright (1975, 1976). This suggestion, which relates to insights encountered in Chapter Two, is pursued in §3.6. In §3.7, it is shown that vagueness *qua* borderline cases (when properly construed so as to exclude terms we are stipulated to give rise to borderline cases) and vagueness *qua* epistemic tolerance are in fact conceptually equivalent, contrary to what might be expected. A puzzle left over from Chapter Two, §2.3, is whether the minimal theory of knowledge entails a realist conception of truth. A similar worry applies to the minimal theory of vagueness, for it looks as if this theory entails that the truth-values of borderline statements are unknowable. However, in §3.8, it is found that while the minimal theory of vagueness, and the minimal margin for error principle **MME**, do indeed entail the existence of undetectable truth, this fact lends no support whatsoever to realism. The principle **MME** is available to everyone. Lastly, in §3.9, I re-configure the anti-internalism argument employed at the end of Chapter Two to show, that there must be higher-order vagueness, contrary to what some have argued.

3.1 Minimalism and vagueness

There are two sorts of minimal theories of vagueness: the sort of theory that is, can, or ought to be endorsed by those who sponsor minimalism concerning truth, and the sort of theory

which endeavours to set forth some uncontroversial characterisation of vagueness with a view to deriving some fertile theorems which all parties to the dispute (whether they like it or not) are beholden to accept. What shape the former sort of theory might be able to take is not the business of this thesis. It is the latter sort of theory, which we may simply call the minimal theory of vagueness, which will occupy us in this chapter. What should we reasonably expect or demand from such a theory?

The traditional goal of any theory of vagueness has broadly been two-fold: to solve the sorites paradox (in all its many guises) and to identify the source of vagueness. We should not expect or demand that a minimal theory of vagueness be able to furnish a generally acceptable solution to the sorites paradox. Were such an uncontroversial solution available that would indeed be gratifying, but no such solution seems in prospect. Likewise, while all parties can agree that much of natural language is vague we should not demand that our minimal theory identify the *source* of this linguistic vagueness—that is the business of some substantive conception. In the light of this, there are a whole range of specific questions which the minimal theory is ill-equipped to answer. Here is just a selection: Is vagueness an exclusively linguistic phenomenon—or is the world in some sense vague? In what exact ways does predicate-vagueness differ from subject-vagueness? Does the existence of vagueness require us to restrict or augment classical semantics and/or first-order logic? If so, what restrictions or additions are required? Must bivalence fail for vague language, for instance? Can we even say that there *is* a logic of natural language, or does logic only apply to an imaginary celestial existence? Is vagueness eliminable? Is a precise meta-language a pipe-dream? Can a convincing model of higher-order vagueness be given? What ramifications does our study of vagueness have for the realism/anti-realism debate, and vice versa?

The inability of the minimal theory to answer such questions might make it appear entirely nugatory; but appearances are misleading. Subsequent discussion will show that this theory is nonetheless equipped to address the following range of questions: Is it possible to isolate the necessary and sufficient conditions for when a sentence counts as vague? Is vagueness at bottom the phenomenon of borderline cases? In what way can we say that vague language is *tolerant*? What is it to say that an expression suffers from higher-order vagueness? Perhaps the minimal theory is rich enough to address many more questions than these; but these are the ones which will occupy us in this chapter. To that end, two immediate qualifications are in order.

Firstly, given that our minimal theory will not illuminate the source of linguistic vagueness, nor furnish a solution to the sorites paradox, then it would seem that this theory

will far from exhaust our understanding of vagueness. As a result, the prospect of giving a *strong* minimal theory of vagueness (one composed of platitudes which exhaust the content of the predicate 'is vague' and cognate expressions) seems ruled out from the outset. However, such pessimism may in the end be unjustified. Indeed, subsequent discussion will show that it proves possible to isolate the constitution of vagueness without solving the sorites paradox and without locating the source of linguistic vagueness. Pending substantiation of that thought, just as with the minimal theory of knowledge developed in the two previous chapters, our minimal account of vagueness will be methodological. Hence, we will assume as a working hypothesis that the content of 'is vague' and cognates can indeed be exhausted by laying down a list of platitudes—in fact, as it turns out, by laying down a neutral, and theoretically lightweight definition of vagueness.

Secondly, it was mentioned in the preamble above that the minimal theory of vagueness should be as neutral as possible not only on philosophical matters but on *logical* matters also. It thus seems that the minimal theory of vagueness can only be developed using some suitably uncontroversial (and thereby very weak) background logic. Since there is little agreement as to the correct logic of vague language, and moreover since there is no general agreement about the correct logic for the language we are entitled to employ in theorising about vagueness (i.e., what we may loosely call the metalanguage), then the project of developing an acceptable minimal theory looks rather bleak. If we take that worry seriously then it looks like there is no scope for the vagueness debate to begin at a mutually agreed point. Every candidate minimal characterisation of vagueness proposed would presuppose some background logical principles which have been, or at least might be, disputed by partisans to the debate as a whole. There would thus be a very real sense in which there would be no genuine disagreement about the character of vague language at all since each partisan would mean something different by the predicate 'is vague'.[2] This (familiar) worry is deep, but not insurmountable.

One way to combat this concern is to adopt what may be termed the *bold* strategy of adopting classical logic as the background logic in which we theorise about particular features of vagueness until such point as such an adoption proves to generate tangible controversy. Williamson (1997a) has recommended a similar bold strategy as the most

---

[2] Cf. Dummett (1991, pp. 68-74) on the problem of debating the revision of logical laws without presupposing such a revision in the language in which the debate is framed. The problem is particularly pressing when a Tarskian-style truth-theory is adopted. As Dummett notes, such a theory will 'leave us without a means to decide what our logic should be [for] the logic that can be shown, by appeal to a theory of truth this kind, to hold in the object-language is directly sensitive to the logical laws assumed to hold in the metalanguage'.

sensible methodology one can adopt when approaching philosophical problems (particularly the problem of vagueness) which might initially seem to demand a revision of classical logic in the object-language or meta-language. Williamson urges that

> one hold's one logic fixed, to discipline one's philosophical thinking [...] in the long-run the results of the discipline will be more satisfying from a philosophical as well as from a logical point of view (*ibid.*, p. 218).

In contrast to Williamson, the suggestion here is not that we should adopt classical logic (in either the metalanguage or object-language) to discipline one's philosophical thinking about *every* aspect of vagueness (including those aspects of vagueness that we must take account of in addressing the sorites or identifying the source of linguistic vagueness). Rather, the suggestion is that we should retain classical logic until such time as this proves to undermine the goal of the minimal theory to furnish an uncontroversial basic characterisation of vagueness. It's a further question whether Williamson's methodology is appropriate to the development of a substantial theory of vagueness (one that solves the sorites and identifies the source of linguistic vagueness). That further question need not worry us here. It is enough that we have a rationale with which to begin our minimal investigations.

How then might we minimally define vagueness? One way in which we might do so is via what Sainsbury has called 'the characteristic sentence' approach. This involves finding

> a sentence schema, containing a schematic predicate position, such that the sentence resulting by replacing the schematic element by a predicate is true if that substitute is a vague predicate (Sainsbury 1991, p. 170).[3]

This seems like a promising start, but we need to adjust this approach in two key respects. Firstly, we must generalise this schema to accommodate the possibility of sentential vagueness which does not necessarily depend on predicate vagueness.[4] Secondly, in order for this approach to offer a constitutive minimal definition of vagueness we must demand that satisfaction of the sentence schema is not merely necessary but also *sufficient* for the vagueness of some sentence. We thus need to find a schema, containing a schematic sentence position or positions, such that the sentence resulting by uniformly replacing the schematic

---

[3] Something like the characteristic sentence approach was first offered by Wright (1987, pp. 282-8). Sainsbury's employment of the characteristic sentence approach is merely provisional as he proceeds to argue that one cannot satisfactorily identify vagueness in this way (Sainsbury 1991, pp.13-5). It is the aim of this chapter to isolate a characteristic sentence which is entailed by almost all conceptions of vagueness, including, I take it, the conception offered in Sainsbury (1990, 1991).

[4] Sainsbury is, of course, well aware of this possibility and has scrutinised non-predicate vagueness, and its possible sources, in some detail in (see Sainsbury 1989b, 1995b).

elements by a particular sentence is true if and only if that substitute is a vague sentence. Can we isolate a sentence schema which would be acceptable to all partisans? One natural place to start is by looking at the property of *being sorites-susceptible*.[5]

3.2 Vagueness qua sorites-susceptibility

Arguably the most general (and least controversial) way to characterise (sentential) vagueness is by reference to the sorites paradox.[6] Say that a declarative sentence is vague just in case this sentence is *sorites-susceptible*. Can we isolate an uncontroversial characteristic sentence which exploits this basic feature of vague expressions?

Suppose we have some sentence S such that the truth of S in a case α depends only on the value $v(\alpha)$ taken by some discretely or continuously varying parameter v in α, where v (let us say) takes non-negative (real or rational) numbers as values. For example, in a simple case, if S is the sentence 'the bath is hot', then v will be the temperature of the water in the bath.[7] Where c is some small positive real number, then it seems initially plausible to say that:

(SS1) $(\forall \alpha)(\forall \beta)$, if $|v(\beta)-v(\alpha)| < c$ then S is true in α if and only if S is true in β

(where α and β range over actual and counterfactual cases). This is to say that where the difference between the value taken by v in α and the value taken by v in β is suitably small, the sentence S will be true in both cases or neither. If the temperature of the bath-water in the α case and the temperature of the bath-water in the β case differs only slightly, then the bath is either hot in both cases or not-hot in both cases. Another way of formulating this claim is

---

[5] In order to keep the discussion as general as possible, in what follows I will for the most part focus on sentential vagueness rather than predicate-vagueness or subject-vagueness.

[6] In what follows, we shall take the primary bearers of vagueness to be declarative sentences, rather than statements or propositions. That may be a controversial step in developing a substantial conception of vagueness (see e.g. Williamson 1994, p.187), but nothing particularly turns on this issue when developing the minimal theory of vagueness. One could equally specify a characteristic sentence schema with propositional variables or letters which schematise utterances which say that something is the case (assertions, suppositions, conjectures, and so on).

[7] Vague terms are typically associated with some dimension or dimensions of comparison. The predicate 'is tall' is one-dimensional (with respect to some comparison class) as it merely governs the dimension of heights. The concept *tall* characteristically takes a positive, a comparative, and a superlative: a person can be tall, taller, and the tallest. The predicate 'is humid' governs (at least) two dimensions: the temperature and water content of air. Colour predicates govern the three dimensions of hue, saturation, and brightness. The predicate 'is hirsute' is multi-dimensional: the thickness, length, colour, texture, distribution, and number of hairs all affect its application.

to say that there are no cases α and β (across which the parameter v varies by some small amount) whereby S is true α and not-S is true in β, which we express as follows:

(SS2) $\sim (\exists\alpha)(\exists\beta)$ such that $|v(\beta)-v(\alpha)| < c$ and S is true in α and not-S is true in β

One immediate suggestion is to employ the schemas **SS1** and **SS2** as characteristic sentences. The idea is that if a given substitution S_1 of S, makes the characteristic sentence **SS1** (or **SS2**) true then the sentence S_1 is vague; conversely, if S_1 is vague then **SS1** (and **SS2**) will be true. Notice that (i) satisfaction of **SS1** or **SS2** leaves it open whether the vagueness of S_1 issues from the predicate or subject terms contained in S_1, or indeed from both types of term, and (ii) any substitution of S must be the sort of sentence whose truth is determined by the degree of variation in one or more graded or continuous parameters $v_1, \ldots v_n$. Any conception of vagueness which can or does constitutively define vagueness via the characteristic sentences **SS1** or **SS2** we may call a minimal conception of vagueness *qua* sorites-susceptibility. On this conception vagueness just is sorites-susceptibility. Have we given a satisfactory minimal definition of vagueness?

The trouble with the schemas **SS1** and **SS2** is that they can both be used to generate paradox (given further uncontroversial assumptions). Let's take each schema in turn. The most familiar sorites template can be given as follows, where we employ the sentence 'A bath of $n°$ is hot', and where premise **A2** (the so-called induction step) is effectively derived from **SS1**:[8]

(A1) A bath of water temperature 100° is hot
(A2) For all n, if a bath of $n°$ is hot then a bath of $n-1°$ is hot
(A3) A bath of water temperature 0° is hot

Call this general form of the sorites the *A*-sorites. Premise **A2** appears to be highly plausible; premise **A1** (the induction base) appears unimpugnable; and the absurd **A3** is derived either by mathematical induction or via one hundred applications of $\forall$-elimination and detachment. All sides can agree that the *A*-sorites represents a logical paradox in that by apparently valid reasoning from apparently sound premises one can derive a patently absurd conclusion. The key premise **A2** codifies the (initially) highly plausible thought that a drop of temperature of one degree cannot make the difference between a hot bath and a bath which is not hot. (Were

one to think that one degree could mark the difference then we need only consider a smaller c-value such as $0.001°$.) Despite the initial plausibility of **A2**, one might nonetheless feel logically obligated to treat this paradox as a *reductio* of **A2**. To do so is to be committed to

(**A4**) There is an n, such that a bath of $n°$ is hot and a bath of $n-1°$ is not hot.

which on the face of it is just to say that 'is hot' is not after all a vague predicate. Our naive intuitions seem to tell us that principles like **A2** are true of vague sentences and false of non-vague sentences. Our naive intuitions thus generate paradox.

The principle **SS2** generates a different form of the sorites paradox (given further uncontroversial assumptions), as follows: our naive intuitions also tell us that

(**B1**) There is no n such that a bath $n°$ is hot and a bath of $n-1°$ is not hot

(where **B1** is derived from **SS2**). It is uncontroversial that a bath of $0°$ is not hot; but let us also suppose for *reductio* that a bath of $1°$ is hot:

(**B2**) A bath of $0°$ is not hot
(**B3**) A bath of $1°$ is hot

which entails, given &-I and ∃-I:

(**B4**) There is an n such that a bath $n°$ is hot and a bath of $n-1°$ is not hot

which contradicts **B1**; and so by negation-introduction we infer:

(**B5**) A bath of $1°$ is not hot.

If we further suppose that

(**B6**) A bath of $2°$ is hot

then by parallel reasoning we can infer that there is an n such that a bath $n°$ is hot and a bath of $n-1°$ is not hot. Contradiction. Reject **B6** to infer that a bath of $2°$ is not hot. One hundred applications of this inference pattern allow us to infer the absurd result that a bath of $100°$ is not hot. Paradox. Call this general form of the sorites the *B*-sorites.[9] Were we to feel logically

---

[8] Strictly speaking, we need the Tarskian schema S is true if and only if p (where p is a translation of S) to make the derivation, but one could equally frame the *A*-sorites in the formal mode of speech.

[9] Wright (1987, p. 261) dubs this form of the paradox the 'No Sharp Boundaries Paradox'.

obligated to treat this as a *reductio* of **B1** then (given classical logic) we would be committed to **A4**, which again just seems on the face of it to rule out the obvious vagueness of the predicate 'is hot'. The *A*-sorites and *B*-sorites are the two main sorites templates which any substantial theory of vagueness must seek to defuse in some appropriate fashion. How one might do this does not concern us in this chapter.[10]

Since **SS1** and **SS2** lead to inconsistency (in many formal systems—including classical logic), these schemas cannot ground an uncontroversial definition of vagueness. If they did, they would forbid the view that vague language is both consistent and subject to classical logic and indeed intuitionistic logic (*cf.* Sainsbury 1991, p.172). In general, very few substantive conceptions of vagueness do indeed sanction either **SS1** or **SS2**—plausible as these principles might initially seem.[11] In response to this worry, one might offer a more anodyne characteristic sentence of the same general form. Perhaps something like the following can be employed to capture the minimal constitution of vagueness:

(SS3)   Pre-theoretically (or, according to our naive intuitions) **SS1** appears to be true

(SS4)   Pre-theoretically (or, according to our naive intuitions) **SS2** appears to be true

A characteristic sentence like **SS3** would, I take it, reflect the fact that the major premise **A2** of the *A*-sorites is apparently true when one is first exposed to this paradox. Likewise for the

---

[10] That said, it would be useful at this stage to have some idea of how one might combat these two puzzles. Like all paradoxes there are three broad options: reject some aspect of the reasoning, reject one (or more) premises; or accept the conclusion but find some way of mitigating the absurdity. Perhaps the most natural immediate response to the *A*-sorites is to seek to preserve **A2** and find some fault with the reasoning. A suggestion to this effect is to reject or restrict mathematical induction as applied to vague statements. Smith (1984) rejects induction; while Ziff (1984) and Stephen Weiss (1976) both suggest restricting this principle. However even if induction were an invalid pattern of inference, the fact that the **A3** can be derived using successive applications of ∀-Elimination and *modus ponens* shows this response to be plainly insufficient. Might we then reject detachment or ∀-Elimination? On the face of it to reject either of these rules seems like a 'desperate remedy' (Dummett 1975, p.303). Desperate it may be, but the validity *modus ponens* has been brought into question for vague discourses. Hyde (1997), for instance suggests that a paraconsistent logic is appropriate for vagueness. In his subvaluational logic, *modus ponens* is invalidated and the *A*-sorites is accordingly defused. It is testimony to the depth of the problem of vagueness that one might seek to dispense with such a basic rule of inference in order to combat the paradox. Are there not less drastic options? One might think that it is better to reject ∀-Elimination but maintain the validity of detachment. Even if such a move were well motivated, it does not address all forms of the sorites paradox. It is well known that one can reconstruct the *A*-sorites as a series of conditionals. This is effectively the form of the sorites which appeared in antiquity. (See Williamson 1994, ch.1.) The paradox then turns on the unwillingness of speakers to locate a conditional in the series which they would be prepared to reject. If each conditional goes unquestioned one can derive the unpalatable **A3** by successive applications of *modus ponens*. So even if ∀-elimination were invalid, this form of the paradox would still be intact. The most popular response to the sorites is to reject the induction step. E.g. Fine (1975) and Keefe (2000) do from within supervaluational logic; Putnam (1983) and Wright (2001) do so from within intuitionistic logic; and Cargile (1969), Campbell (1974), Sorenson (1988), and Williamson (1992b, 1994) do from within classical logic and classical semantics. (See the rest of this chapter and the next for some relevant discussion.)

key premise **B1** of the *B*-sorites. To be sorites-susceptible is not to be committed to absurdity *per se* but is rather to be *apparently* subject to soundness of the sorites paradox. On this basis, say that a sentence is vague just in case it is sorites-susceptible in the sense just given, just in case when substituted into **SS3** and **SS4** it renders these schemas true.[12] Surely this ought to be agreeable to all.[13]

While the characteristic sentences **SS3** and **SS4** are, presumably, uncontroversial, they are also unspecific. How are we to give a rigorous account as to what is meant by the qualifiers 'pre-theoretically', 'according to our naive intuitions' or 'appears to be'? One person's pre-theory may be so coarse that such principles like **SS1** and **SS2** simply strike them as gibberish. On another's pre-theory perhaps these schemas appear to be obviously false. Moreover, we also need an account of why it is that our naive intuitions incline us to accept **SS1** and **SS2** (see §3.7). Perhaps some deeper feature of vague expressions explains that inclination in which case vagueness defined in terms of sorites-susceptibility is not a particularly informative characterisation. So while a definition employing **SS3** and **SS4** may record a genuine conceptual (if rather unspecific) insight, and while it goes some way to ensuring that partisans to the vagueness debate are not talking past each other, it does not seem to record an *explanatory* insight. There is a strong sense in which a sentence is sorites-susceptible *because* it is vague, and not vice versa. Sorites-susceptibility is secondary in the explanatory order. We should look elsewhere for our minimal definition of vagueness.

3.3 Minimal vagueness qua borderline cases: the indeterminist conception

On what is perhaps the most prevalent conception, vagueness is the phenomenon of borderline cases. This conception is so widespread that Sorenson (1985, pp. 135-6) can confidently say that:

> Although there is considerable disagreement over the nature of the defectiveness and exact nature of vagueness, there is general agreement that predicates which possess borderline cases are vague predicates.[14]

---

[11] The conceptions of vagueness offered by Dummett (1975), Unger (1979), Wheeler (1979), and Hyde (1997) all sanction **SS1**.

[12] Where again, any substitution of *S* must be the sort of sentence whose truth is determined by the degree of variation in one or more graded or continuous parameters $v_1, \ldots v_n$. (This qualification will often be left inexplicit in the rest of this chapter.)

[13] Sorenson (1985), for instance, takes sorites-susceptibility to be one hallmark of the vague. See also Keefe and Smith (1996, p. 3).

[14] The tradition of defining vagueness primarily in terms of borderline cases dates back to Peirce (1902, p. 748), was continued by Black (1937, p. 30), and receives its fullest expression in Fine (1975).

How then might we define the relevant notion of borderline case from within a conception which is as neutral as possible on matters both logical and philosophical?

Wright (1987, p. 262) has offered the thought that 'when dealing with vague expressions, it is essential to have the expressive resources afforded by an operator expressing definiteness or determinacy'. If this claim is correct then it will hold just as much for the minimal theory of vagueness we are trying to articulate here as it will for any further substantive conception of vagueness. So in this section, taking our lead from Wright, let us pursue the provisional strategy of defining a notion of vagueness *qua* borderline cases via some (as yet unspecified) notion of definiteness or determinacy. But which notion is to be employed—definiteness or determinacy? It will do no harm to assume, following Williamson (1996b, p.44), that definiteness and determinacy, in the context of vagueness at least, are fully interchangeable notions.

A first promising candidate characteristic sentence for the indeterminist minimal theory might be given as follows:

(DT1)   ($\exists \alpha$) In α, it is not determinately the case that S is true and it is not determinately the case that not-S is true

Again, the idea is that if a given substitution S_1 of the schematic condition S, makes the characteristic sentence **DT1** true then the sentence S_1 is vague; conversely, if S_1 is vague then **DT1** will be true.[15] Any conception of vagueness which can or does define vagueness via the characteristic sentence **DT1** we may call an *indeterminist* minimal conception of vagueness *qua* borderline cases. Such a conception is intended to capture the thought that a sentence is vague just in case it takes a status intermediate between determinate truth and determinate non-truth (falsity). Is this characterisation defensible?

Satisfaction of **DT1** cannot be sufficient for the presence of vagueness. A familiar complaint in this regard is that it is a mistake to take borderline cases *per se* to be constitutive of vagueness.[16] To illustrate, suppose we stipulate that the open sentence 'x is an oldster' is

---

[15] The fact that 'α' ranges over both actual and counterfactual situations allows us to capture Fine's distinction between intensional and extensional vagueness (Fine 1975, p.266). A sentence is extensionally vague just in case it does give rise to borderline cases (given the way the actual world is) and is intensionally vague just in case it *could* give rise to borderline cases. The sentence 'Timothy Williamson is thin' is extensionally vague (as he concedes) and remains intensionally vague in situations where all people are either determinately thin or determinately not thin.

[16] See Dummett (1973, pp. 646-7); Wright (1975, p. 329); Sainsbury (1989a, p. 34-5, 1991, pp. 173); Hyde (1994, pp. 35-6).

(determinately) true of every person over sixty-eight years of age, (determinately) false of those persons under sixty-five years of age, and neither (determinately) true nor (determinately) false of the remainder. If a speaker applies this term to persons who are between sixty-five and sixty-eight then we are entitled to say that they have done something not quite right and done something not quite wrong according to the dictates of the stipulation. But since 'x is an oldster' is neither determinately true nor determinately false of the intermediate cases then given **DT1** it counts as vague, even though intuitively we should be inclined to say that the term is *not* vague but rather, in some sense, semantically incomplete. This species of indeterminacy *per se* is not vagueness, since the term 'oldster' draws a perfectly sharp and clearly identifiable three-fold division across its associated dimension of comparison.

In reply to this problem, it might be said that the problem of the sentence 'x is an oldster' stems in essence from failing to accommodate the possibility of higher-order vagueness in setting forth our characteristic sentence. What is meant by higher-order vagueness? Very roughly, say that a sentence is higher-order vague just in case it not only gives rise to borderline cases (cases where it is neither determinately true nor determinately false) but borderline cases to those borderline cases (cases where it is neither determinately true nor determinately false that the sentence is neither determinately true nor determinately false), and borderline cases to these borderline cases, and so on. The sentence 'x is an oldster' would count as genuinely vague if it were also to give rise to borderline cases to the borderline cases, and in turn borderline cases to those borderline cases, and so on. Since it does not, it is not genuinely vague. But can we rehabilitate a constitutive indeterminist minimal account *qua* borderline cases by appealing to some form of higher-order vagueness without at this stage giving an explicit (and perhaps controversial) model of higher-order vagueness?

One way to do this is to borrow the strategy of Hyde (1994) who has offered the thought that the notion of 'borderline case' is ambiguous between the type of borderline cases that stems from such terms as 'oldster', and genuine borderline cases of vagueness where higher-order vagueness 'is built in from the very start' (*ibid.*, p. 40). How then might we adjust the characteristic sentence **DT1** to accommodate Hyde's requirement that vagueness *qua* borderline cases automatically ensures that radical higher-order vagueness is built in from the outset? Hyde (1994, p. 39) in effect suggests that one need not make an *explicit* reference to the existence of higher-order borderline cases in our characterisation of vagueness at least

insofar as we ensure that we have distinguished the type of indeterminacy that is constitutive of vagueness (call it **indeterminacy**) and the type of indeterminacy (just call it indeterminacy) that is characteristic of such terms as 'oldster'. For Hyde, there is no real problem of higher-order vagueness; but rather the

> problem surrounding higher orders of vagueness arises when one tries explicitly to state something about the nature of vagueness that manifests itself in the characterisation anyway—the phenomenon of higher-order vagueness. [...] There are border border cases for vague predicates, but this need not be stated as part of the analysis of the concept of predicate-vagueness [...] One is simply repeating oneself and adding nothing new (*ibid.*, p. 40).

So while Hyde accepts that higher-order vagueness is genuine feature of vague language (see p. 40), he rejects any suggestion that one can *explain* what vagueness *qua* borderline cases amounts to by reference to the thesis that a vague expression gives rise to borderline cases, and borderline cases of those borderline cases, and so on (or indeed by reference to any more rigorous statement of higher-order vagueness). It's not that Hyde disallows us from expressing what is meant by higher-order vagueness in this way, it's rather that in doing so one adds nothing to our understanding of vagueness: in the order of explanatory priorities, our grasp of this thesis (or our grasp of a more rigorous formulation of higher-order vagueness) is secondary to our grasp of the basic notion of **indeterminacy**.

If this is right then one can simply side-step the problem of the sentence 'x is an oldster' by offering the following characteristic sentence:

(**DT2**) ($\exists \alpha$) In α, it is not **determinately** the case that S is true and it is not **determinately** the case that not-S is true

Again, a sentential substitution is vague just in case **DT2** is true for that substitution.

Should all partisans to the dispute accept **DT2**? An immediate worry with **DT2** is that it is not yet a settled question whether any respectable theory of vagueness should indeed entail that vague terms are higher-order vague—as the requisite notion of **indeterminacy** demands. Since not all theories of vagueness entail the existence of higher-order borderline cases (not even implicitly) then it looks as if we have gained a sufficient condition for the indeterminist minimal definition at the expense of losing a necessary one—and with it we appear to have lost the promise of ensuring that the dialectic of the debate can begin at a mutually agreed

point.[17] The tempting reply is to say that a theory which doesn't recognise higher-order vagueness is just obviously false. Given the resources afforded by the minimal theory proposed thus far we are not yet in a position to settle this matter. But even if we ought to recognise the existence of higher-order vagueness (*qua* borderline cases), and even if we are willing to accede to Hyde's thesis of the ambiguity of 'determinately', one might nonetheless worry that the notion of determinacy is as yet too unspecific to feature in our minimal theory of vagueness. Is this a *sui generis* notion or can we explicate its nature in some substantial way? Is it an entirely non-epistemic notion? Indeed, do we really need a notion of determinacy at all in giving a (minimal) characterisation of vagueness (as Hyde, following Wright 1987, thinks)? To put **DET2** to use in our minimal theory we must first address these questions. (Since it's as yet an open question whether 'determinately' or 'definiteness' are ambiguous, in what follows I shall simply employ these terms in non-bold format.)

3.4 Determinacy and definiteness: a brief survey

One worry one might have with **DET2** (and indeed **DET1**) is that natural vague language does not in fact contain the requisite predicates 'is determinately true/false' or their material-mode counterparts 'It is determinately true/false that'. Moreover, if natural language were appropriately extended to include these operators, then the intuitions of natural language speakers could not be reliably employed to determine whether the appropriate instances of some characteristic sentence were true or not (see Sainsbury 1991, p.174). This seems correct; but in reply it might be said that the intuitions of native speakers are not at issue when one is attempting an *explication* (i.e. a logical reconstruction) rather than some mere (descriptive) analysis of the phenomenon of vagueness. So, the reply runs, there ought to be no principled obstacle to specifying an extended language in which a notion of definiteness or determinacy is suitably introduced (*cf.* Williamson 1999, p. 129, fn.2). (Perhaps indeed such a regimentation would appropriately disambiguate 'determinately' along lines suggested by Hyde.) For the purposes of isolating a minimal theory of vagueness, how might we *explicate* the relevant notion of determinacy (and its cognate notion of definiteness)?

---

[17] Wright (1987, 1992b), Kamp (1981) and Sainsbury (1990,1991) amongst others, have doubted the existence of (non-terminating) higher-order vagueness. See §3.9.

Dummett (1978, p. 256) has said that 'in connection with vague statements, the only possible meaning we could give to the word "true" is that of "definitely true"'. Likewise, for 'false' and 'definitely false'. This suggests that—contra Wright (1987)—we can do without talk of definiteness and determinacy in our characterisation of vagueness, for if Dummett is correct, to say that a statement is (extensionally) vague is really to say no more than that it is neither true nor false. Dummett nonetheless maintains that we should not dispense with the 'determinately' operator for the following reasons: the notion of truth which is relevant to vague statements is a non-distributive notion (see his 1991, pp.75-6). In particular, Dummett argues that a disjunction 'x is either orange or red' can be true even though x is on the red-orange borderline such that neither disjunct is true (which is not to say that one or more disjuncts is false). But how can we mark the difference between non-distributive true disjunctions and distributive true disjunctions? Enter the adverb 'determinately' (or 'definitely'). This adverb has, for Dummett, a special *force*—it can be used to record the fact that a disjunction is not only true, but that it is true *in virtue of the fact* that at least one of its disjuncts is true. It is for this reason that Dummett urges we should always formulate the principle of bivalence as saying that every statement is *determinately* either true nor false, so as to rule out the possibility that a class of statements are all either true or false, but that it is not (determinately) true which. Whatever the merits of this proposal, it is clear that this analysis is not compatible with any conception of vagueness in which truth does distribute over disjunctions. In general, we should demand that the minimal theory of vagueness must not exclude from the outset that the logic and semantics of vagueness is classical. To do that is to rule out the possibility that vagueness is an entirely epistemic phenomenon which demands no restriction of classical semantics or classical logic. There is no hope that a Dummettian conception can illuminate the sense of definitely we require for **DT2**.

The point generalises. From the minimal perspective, it is illegitimate to confer a non-epistemic interpretation to the adverb 'determinately'—tempting as that construal may be. Determinate truth, on the minimal conception of vagueness, cannot be taken to mean *truth to degree 1* (as a many-valued theorist might argue), or *truth under all admissible sharpenings* (as Dummett would think), for again we must allow that vagueness might be, after all, a special species of ignorance. As Wright has said, it cannot be a basic datum that indeterminacy is non-epistemic phenomenon, for this is just to saddle ourselves from the outset with a 'proto-theory' of vagueness (Wright 1995, pp. 133-4; see also Horwich 1997, pp. 929-30).

71

The point also applies to those who have offered what we might term *quasi-semantic* interpretations of determinacy or definiteness. McGee and McLaughlin (1995, p. 209) suggest that 'to say that an object *a* is definitely an *F* means that the thoughts and practices of speakers of the language determine conditions of application for the word *F*, and the facts about *a* determine that these conditions are met'. Statements are vague, accordingly, when the thoughts and practices of speakers in some sense under-determine what their conditions of correct application are. This might of course in the end be the correct view of vagueness, it is simply that it is incompatible with the (standard) epistemic conception of vague language whereby the thoughts and practices of speakers do fully determine the conditions under which vague terms are true or false—it is just that in borderline cases we are unable to tell whether or not these conditions obtain.

Perhaps, then, determinacy (or definiteness) is a *sui generis* notion? Field (1994b, p. 111, 2001, p. 227) has offered the view that 'definitely' is a primitive expression whose meaning is to be grasped in the same way in which speakers might be said to grasp the meaning of the standard logical operators—via their introduction and elimination rules.[18] On this model, we can *implicitly* define what determinacy is by the operational rules for the 'definitely' operator, and then on that basis offer a constitutive definition of vagueness via some appropriate characteristic sentence. Field's conception looks to be of little use to the minimalist. A *sui generis* conception of definiteness precludes the possibility that 'definitely' or 'determinately' can be explicated in terms of such epistemic notions as knowledge, clarity, or knowability. Again, since we do not want our minimal theory to represent a proto-substantive theory of vagueness, one which disqualifies the epistemic conception from the beginning, then Field's conception of determinacy can form no part of the minimal theory.

Might there then be an *epistemic* reading of 'determinately' which is compatible with all conceptions of vagueness? Wright (1995, pp. 144-6) has suggested that 'determinately' might best receive a *quasi-epistemic* reading: roughly, when a statement *P* is determinately true then for a speaker *s* to judge that not-*P* means that *s*'s verdict is 'cognitively misbegotten'—the lighting might be bad, *s* might be drunk, tired, distracted, or forgetful. In borderline cases, cases where *P* is neither determinately true nor determinately false, Wright envisages that there can be 'faultlessly generated—cognitively un-misbegotten—conflict': subjects may permissibly disagree about the borderline cases, where the notion of permissible

---

[18] Though, Hyde does not offer the thought that the meaning of **determinacy** is given by the introduction and elimination rules governing the **'determinately'** operator, he also seems to be sponsoring a *sui generis* reading of **determinacy**.

disagreement is, for Wright, of the very essence of vagueness (Wright 1987, p. 277, 1995, p. 138). Moreover, Wright urges that this interpretation is compatible with both standard epistemic and standard non-epistemic conceptions of what it is to be a borderline case.[19] While that may be so, Wright's quasi-epistemic reading of 'determinately' requires the claim that the thesis of permissible disagreement is of the very essence of vagueness. Yet the thesis of permissible disagreement has proved hard to stabilise (see Wright 2000, pp. 55-62 for the most relevant evaluation). Even if it is stable, it's not at all clear that one can stabilise it given uncontroversial resources. Consequently, Wright's quasi-epistemic reading can form no part of the minimal theory.

Might 'determinately' receive a more overtly epistemic reading? Williamson has famously argued that it is not just possible to give 'definitely' or 'determinately' some (overt) epistemic reading, it is the *only* illuminating and coherent reading we can give to these adverbs in the contexts of vagueness (Williamson 1994, pp. 194-5; and especially his 1995). He suggests that 'definitely' may in effect mean something like 'knowably'. The trouble with this suggestion is that Williamson appears to take it to be a basic datum—on any conception of vagueness—that the truth-values of borderline statements are unknowable (either because they essentially defeat our powers of discrimination or because there is no fact of the matter to be known in borderline cases). This is not a basic datum. On certain hitherto neglected conceptions of vagueness—conceptions which we will touch upon in §3.8—the truth-values of borderline statements are unknown but in principle knowable (via some method or other). A minimal epistemic reading of 'determinately' must not entail that the truth-values of vague sentences are verification transcendent, for this is just to preclude these neglected conceptions from the start.

It now begins to look that that the only way in which the dialectic of the vagueness debate can begin at a mutually agreed point is to leave the requisite notion of determinacy or definiteness employed in some characteristic sentence such as **DT2** unspecified. Each substantive conception of vagueness would then sanction the minimal indeterminist

---

[19] To elaborate: in the epistemic case, suppose P is true, but unknowable in borderline cases, then a (non-inferential) verdict that not-P is *not* cognitively misbegotten, according to Wright, as there is no sense in which one can blame the speaker for their mistaken verdict when the truth of P is undetectable. On an indeterminist conception of vagueness where the adverb 'determinately' is non-epistemic and not strictly redundant (e.g. on Field's view, but not Dummett's) an object a is borderline for a predicate F just in case a is neither determinately F nor determinately not-F. Thus, the thought goes it would be not be quite right and not be quite wrong to assert that a is F (likewise for an assertion that a is not F) for the matter is unsettled. And so, Wright envisages that subjects can be represented to permissibly differ in their verdicts on such a conception (or at the very least that such a conception is compatible with the thesis of permissible disagreement). In his (2000), Wright dispenses with any commitment to permissible disagreements.

conception only insofar as it is permitted to interpret 'determinately' according to considerations local to that conception. If this were so, then Wright's dictum that it is essential to have the expressive resources afforded by some notion of definiteness or determinacy begins to look of little use in specifying the sort of minimal theory we would ideally like to have. It is theoretically unsatisfactory to allow the minimal indeterminist conception to command universal assent simply in virtue of employing a multiply ambiguous conception of determinacy—for in that case we can hardly be said to have given a minimal *definition* of vagueness at all.

One (perhaps obvious) option remains: might we simply take 'definitely' or 'determinately' to mean 'clearly' or 'known to be'? In so doing, for one thing, there is no implication that truth is potentially verification-transcendent. Arguably, every conception of vagueness entails that a vague declarative sentence is neither clearly true nor clearly false, that if an object a is a borderline case for some predicate F then neither F nor not-F are clearly true of a. Furthermore, one advantage of taking 'definitely' to mean 'clearly' or 'known' is that the logic of clarity/knowledge, in comparison to the logic of indeterminacy, commands a greater (though far from absolute) consensus. (Perhaps indeed this proposal would then enable us to make headway on the question as to whether or not there is higher-order vagueness.) Lastly, this suggestion would also have the advantage that we can simply dispense altogether with the any notion of definiteness and determinacy in giving our minimal theory of vagueness, and in so doing, we can rid the minimal theory of the misleading non-epistemic overtones that the notion of determinacy inevitably carries. It thus looks like we must leave behind the characteristic sentence **DT2**, and instead endeavour to specify how a notion of clarity can feature in a minimal characterisation of vagueness,

3.5 Minimal vagueness qua borderline cases: the epistemic conception.

It is not too far wrong to say that there is an emergent consensus that the proper minimal theory of vagueness ought to be stated in epistemic rather than non-epistemic terms. In particular, that one can usefully employ a notion of ignorance or unclarity to ground an uncontroversial definition of vagueness. (Though it has to be said that those make this claim are not explicitly interested in developing a minimal theory of vagueness.) Sainsbury, for one, has urged that:

74

All theorists can agree that a certain kind of ignorance is a sign of vagueness. We do not know whether or not some people are tall, not because we do not know how tall they are, but because we do not know whether being that tall counts as being tall [...] We do not know whether we are still on Snowdon, not because we do not know where we are (we might know our map reference, or our precise distance from the summit) but because we do not know whether being here counts as being on Snowdon. Let us call cases which do or would give rise to such ignorance borderline cases (1995b, p. 64).

Likewise, Williamson (1997a, p. 921) agrees that vagueness 'is the phenomenon of borderline cases', and that we can at least *ostensively* define what it is to be a borderline case by giving examples:

At some times, it was unclear whether Rembrandt was old. He was neither clearly old nor clearly not old. The unclarity resulted from vagueness in the statement that Rembrandt was old. We can even use such examples to define the notion of vagueness. An expression or concept is vague if and only if it can result in unclarity of the kind just exemplified. Such a definition does not intend to display the underlying nature of the phenomenon. In particular, it does not specify whether the unclarity results from the failure of the statement to be true or false, or simply from our inability to find out which. The definition is neutral on such points of theory (1994, p. 2; see also p. 202).[20]

Any conception of vagueness which sanctions such a characterisation we may call an *epistemic* minimal conception of vagueness *qua* borderline cases. Though it is tempting to read the above remarks as providing the materials for a minimal definition of vagueness given in terms of necessary and sufficient conditions, Sainsbury, Williamson, and Tye all resist offering such a definition. Is such resistance justified? This will be the question which will preoccupy us for the rest of this section.

Suppose we now offer the following characteristic sentence:

(CL1)   ($\exists \alpha$) In α, it is not clearly the case that S is true and it is not clearly the case that not-S is true

where we not only require that an acceptable substitution of S must be the sort of sentence whose truth is determined by the degree of variation in one or more graded or continuous underlying parameters $v_1, \dots v_n$, but, crucially, that the source of unclarity must issue from features of the substituted sentence (or from features of the *use* of that sentence) and not from any ignorance as to the underlying v-facts. This is to say that even when a speaker is apprised

75

of all the relevant v-facts, it will still remain unclear whether or not the sentence S is true.[21] (For the time being, we may assume, following Williamson (see his 1994, p.16), that clarity and knowledge (and unclarity and ignorance) are freely interchangeable notions. This is to say, the operators 'it is clearly the case that' and 'it is known that' are fully interchangeable.) Is the constitutive definition using **CL1** at all compelling?

In §3.3, we saw that the possibility of such artificial terms as 'oldster' ruled out using **DET1** to constitutively define vagueness. Effectively the same worry rules out the use of **CL1**. The term 'oldster' is neither clearly true nor clearly false of the intermediate cases (since it is neither determinately true nor determinately false of those cases), and since this unclarity issues from features of the sentence 'x is an oldster' (and not from any ignorance regarding the underlying v-facts), then it counts as vague given the characteristic sentence **CL1**. Again, we should be strongly inclined to say that the term is not vague since it draws a perfectly sharp and clearly identifiable three-fold division across its associated dimension of comparison. Hence, a constitutive minimal definition invoking **CL1** does not allow us to distinguish between vagueness and various distinct but superficially similar phenomena, such as semantic incompleteness.[22] But might one rehabilitate **CL1** by appealing to some form of higher-order vagueness?

---

[20] Tye (1995, p.1) offers a similar neutral characterisation. Keefe and Smith (1996, p.2) follow suit, but then proceed to find it immediately plausible that our unclarity in borderline cases is due to there being no fact of the matter to be clear about.

[21] What sort of relationship might the truth or falsity of S bear to the v-facts? The minimal theory should have some answer to this question. It seems clear that for some possible fact about the truth or falsity of S to actually obtain (call these the S-facts) then some possible v-fact or facts must actually obtain. More specifically, a S-fact will obtain *because* certain possible v-facts obtain, and whatever v-facts obtain will *determine* whether or not S is true. This is just to say that the S-facts (strongly) *supervene* upon the v-facts; that is:

(**SUP**) For any two objects x and y, and for any possible world w, if x has exactly the same v-properties as y (in w) then x has exactly the same S-properties as y (in w).

(where the S-properties are the properties of truth and falsity). It follows from **SUP** that if, in any world w, an object x and an object y diverge in their S-properties then they must also diverge in their v-properties. The converse implication does not necessarily follow (at least for multi-dimensional predicates): Billy and Barney may both be fat (to exactly the same degree) and yet they may diverge quite considerably in their relevant v-properties. If we allow ourselves the convenient simplification that the vocabulary used to represent the v-facts is vagueness-free, then we may simply speak of the v-facts as being the *precise* facts and the S-facts as being the *vague* facts. Accordingly, our question has been answered: the vague facts supervene upon the precise facts. Matters however are not quite so straightforward. We also need to know what kind of necessity is involved is **SUP**. Williamson (1992b, p. 153; 1994, pp. 203-4; 1996d pp. 331-3) has assumed that the type of necessity in play in **SUP** is *metaphysical* necessity—there could not be, as a metaphysical matter, two situations which differ in their F-properties but not their v-properties. One might wish to advance the further, and stronger, claim that it is rather *conceptual* necessity that is in play in **SUP**. A speaker would make a conceptual error in supposing that Billy and Barney might differ in their degree of fatness and yet fail to differ with respect to their v-properties.

[22] One suspects it may have been partly for this reason that Sainsbury, Williamson, and Tye, all resist giving a constitutive definition in terms of unclarity or ignorance.

We saw above that one way in which we can incorporate a notion a notion of higher-order vagueness into **DET1** is to draw on Hyde's thesis that the notion of 'borderline case' is ambiguous between the type of borderline cases that stem from such terms as 'oldster' and genuine borderline cases where higher-order vagueness 'is built in from the very start'. But now we are dealing with an epistemic notion of a borderline case. Hyde's purported ambiguity in the notion of borderline case is 'ultimately as a result of the ambiguity of "determinately"', where Hyde appears to give this operator a non-epistemic reading (see Hyde 1994, p.40). Is the adverb 'clearly' systematically ambiguous in a similar way? If it were then we could offer the following characteristic sentence:

(**CL2**) ($\exists \alpha$) In α, it is not **clearly** the case that S is true and it is not **clearly** the case that not-S is true

The hope is that while the sentence 'x is an oldster' would not satisfy **CL2**, a sentence such as 'x is red', for example, would satisfy **CL2** (where x is on the red-orange borderline).

It is not obvious that 'clearly' is ambiguous in the manner in which 'determinately' might be. On our ordinary understanding of clarity, there is no manifest implication that higher-order unclarity is built in from the very start. Of course, it is open for one to *explicate* that '**clearly**' is to be understood in the requisite way, but to do so is problematic. How else can we fix the truth-conditions of **CL2** without explicitly adverting to the fact that there are (epistemic) border cases, and (epistemic) border cases of those of border cases, and so on? In this case, our grasp of **clarity** would be secondary to our grasp of higher-order vagueness (epistemically construed in terms of clarity). This is bad news for Hyde, but not necessarily bad news for the characteristic sentence approach. It is still open for us to make an explicit reference to (epistemic) higher-order vagueness when setting forth some appropriate characteristic sentence. But, since the existence of higher-order vagueness (*qua* borderline cases) is a matter of controversy (see Kamp 1981; Wright 1987, 1992b; Sainsbury 1990, 1991; Burgess 1990, 1998; Koons 1994) it is idle at this point to do so until such time as we have reassured ourselves as to the existence of higher-order vagueness (see §3.9).

This leaves us with two main options: Follow Williamson, Sainsbury, and Tye, and rest content with an ostensive minimal definition of vagueness *qua* (epistemic) borderline cases and thereby concede that the constitution of vagueness can only be identified from within some substantive conception (if at all) or, rehabilitate the characteristic sentence approach by appealing not to the notion of a borderline case in the first instance, but by reference to some

other salient (and perhaps deeper) feature of vague expressions. We saw in Chapter One that the minimal theory of knowledge was ill-equipped to provide a constitutive definition of knowledge. Indeed given the general difficulty of constitutively defining any philosophically interesting concept, perhaps the promises of the minimal theory have been overstated, and thus we should rest content with the first option. But such pessimism is unjustified. It is theoretically unsatisfactory that our minimal theory of vagueness should just rest content with defining vagueness via exemplars of vague language. In the next section, I hope to show that we can indeed offer a constitutive definition by reference to a further feature of vague expressions. Namely, the phenomenon of what I call *epistemic tolerance*.

3.6 Vagueness qua epistemic tolerance

For some, vagueness is the phenomenon of borderline cases, and there's an end on't. Fine, for instance, takes vagueness to be, in essence, a one-dimensional phenomenon. Bluntly: 'a predicate is extensionally vague if it has borderline cases' (1975, p. 266). We have already seen the shortcomings of such a characterisation given the possibility of such terms as 'oldster'. What is surprising is that Fine makes no reference to a feature of vague expressions which is *prima facie* far more basic—namely, that such expressions draw no *given* boundary across their range of signification.[23] This feature reflects the basic phenomenological datum that along some smooth or graded dimension of comparison governed by some predicate F subjects characteristically do not cognise any boundary between the F's and the not-F's. In this section we will focus on this *prima facie* more primitive dimension to vague language with an eye to articulating some generally acceptable characteristic sentence.

Just what is meant by the phrase 'no given boundary' in this context? There are at least two readings we can give this phrase: to say that vague terms draw no *given* boundary is just to say (i) vague terms draw no *apparent* boundary across their range of signification, (ii) vague terms draw no *clear/known* boundary across their range of signification. Hyde (1994, p.34) has effectively offered (i) as a neutral characterisation of vagueness, while Burgess (1998, p.233) has advanced (ii) in a similar spirit. I propose to follow Burgess rather than Hyde, and employ the factive operator 'It is clearly the case that' in the development of the

---

[23] When reading Fine (1975) one can be left with the disconcerting impression that he is not really talking about vagueness at all. Hence, the pressing need for an articulation of the conceptual relationship between vagueness *qua* borderline cases and this *prima facie* more basic feature of vague expressions. See §3.7.

minimal theory of vagueness. This is a provisional position. There may be value in employing other non-factive operators to express our minimal commitments, but I shall not do so here.[24]

It's worth quoting Burgess in full as it will help elucidate what follows:

> Regardless of the theory of vagueness we adhere to, we all agree that no facts, known or practically knowable, suffice to determine the location of precise boundaries for vague concepts. According to the epistemic theory of vagueness, this ignorance is *entirely* an epistemic matter—vague concepts have sharp boundaries but we can never know their exact locations. Opposed to epistemicism is a view—or a family of views—I shall call indeterminism. The indeterminist agrees with the epistemicist that we lack knowledge of the locations of sharp boundaries to vague concepts but holds that this ignorance has a somewhat more dramatic source—vague concepts have no sharp boundaries for us to be ignorant of (*ibid.*, p. 23).

Let me make two brief comments on Burgess's remarks: Firstly, let's follow Burgess's lead from now on and ensure that the term 'indeterminism' denotes only *non-epistemic* conceptions of vagueness in order to avoid confusion with those conceptions which seek to advance an epistemic or quasi-epistemic reading of 'definitely' or 'determinately'. Secondly, as briefly mentioned above, not every epistemic view of vagueness entails that sharp cut-offs are unknowable. (Though, we have to tread with care here for I will argue in §3.8 that there are different modes of unknowability.) So, from the neutral point of view, we should merely say that speakers lack knowledge of the location of sharp boundaries and not that speakers can never be in a position to know the location of such cut-offs. We have to make room for what I term *optimistic* conceptions of vagueness whereby vague terms draw sharp but in principle knowable boundaries (see §3.8). These caveats aside, Burgess's observations can form the foundation of a minimal characterisation of vagueness that promises to be an improvement on any view which attempts to identify vagueness (solely) by reference to borderline cases.

Arguably, just about every substantive conception of vagueness will—implicitly or explicitly—entail that vague terms draw no clear/known boundary across their range of signification. Henceforth, I shall simply read this as saying that vague expressions draw no clear boundary across their range of signification. More loosely, and in the parlance of Fine, we can read this as saying that vague terms have *blurred boundaries*. It's tempting to read this as substantiating the further thesis that vague expressions draw no sharp boundary across

---

[24] One might want to use the operator 'I am subjectively certain that' or some other barely normative operator. These are suggestions for further work.

their range of signification. This tempting conclusion is not one which is available to the minimal theory of vagueness.[25] The fact that speakers are unable to locate a sharp cut-off between, *red* and *not-red*, for instance, in no way entails that there is no sharp cut-off—at least from the neutral perspective of the minimal theory. The minimal conception must allow that perhaps, after all, there is such a sharp cut-off but we are simply unable to determine its whereabouts (in conceptual space). This point is analogous to the one made above in §3.3; namely, that we do not wish to saddle ourselves from the outset with a proto-theory of vagueness, one which excludes the epistemic conception of vagueness and the concomitant commitment to sharp boundaries.[26]

Since all partisans to the dispute can or do, for whatever reason, accept that vague terms draw no clear boundary across their associated dimension of comparison, it seems right that the minimal conception can and should countenance a further dimension to vagueness: there is vagueness *qua* sorites-susceptibility, vagueness *qua* borderline cases, and vagueness *qua* lack of clear boundaries. In the next section, we shall look at the conceptual connections between the latter two dimensions, for present purposes we need to find a characteristic sentence which appropriately unpacks the claim that vague expressions have blurred or unclear boundaries. To give a more rigorous characterisation we must dispense with talk of 'cut-offs' and 'boundaries' and offer instead a constitutive definition in terms of a what I call *epistemic tolerance*. What is meant by tolerance and epistemic tolerance in this context?

In his (1975, p. 334), Wright provisionally suggested that vague predicates are 'tolerant'.[27] Suppose we have a (monadic) predicate F which governs some dimension of comparison $\varnothing$, then according to Wright

---

[25] Keefe and Smith (1996, pp. 2-3), in their introductory characterisation of vagueness offer the view that 'vague predicates [...] apparently lack well-defined extensions', but then add that on 'a scale of heights, there is *no sharp boundary* between the tall people and the rest' (see also Keefe 2000, Ch.1). From the introductory perspective, this is highly misleading. One should take care that the expression 'sharp' does not carry any epistemic overtones such that no *sharp* boundary comes to mean something like no *clear* boundary. Once 'sharp' is free of epistemic connotation, then the expressions 'no sharp boundary' and 'no boundary' should be thought of as equivalent—there is, strictly speaking, no such thing as an unsharp boundary (*cf.* Sainsbury 1990, 1991).

[26] These observations also bring out why the minimal theory is dialectically valuable. While many theorists not only take it to be a basic datum that vague terms fail to draw sharp boundaries (with the implication that any epistemic or quasi-epistemic conception of vagueness is obviously misplaced), they also take it to be a basic datum that the truth-values of borderline statements are in principle unknowable. The minimal theory ought to be free of such unhelpful prepossessions.

[27] Since the discussion of tolerance in the literature has focussed on predicate-vagueness, I shall follow suit. It should be noted that Wright (in this early paper at least) takes vague terms to be tolerant if a particular view of the language capacity is adopted—a view he calls the 'governing view'. Very roughly, the governing view is the view that to master language is just to internalise a set of rules which give a complete set of instructions as to how the language is to be combined (syntactic rules) and applied (semantic rules). Wright rejects this conception in favour of language which is not rule-governed in this way and which must give priority to behavioural data.

> F is tolerant with respect to $\emptyset$ if [and only if] there is some positive degree of change in respect of $\emptyset$ insufficient ever to affect the justice with which F is applied to a particular case.

A thesis of strong tolerance (**ST**) says that every vague predicate is tolerant in this respect. In general, the degree of change will be depend on the context. A predicate F can be tolerant even when the value of the degree of change can be such as to effect a discriminable difference between cases along $\emptyset$. For large degrees of change, there typically will be a difference as to whether F can be applied to a particular case.[28] Say that a predicate F fails to draw a sharp boundary when F is tolerant in the sense just given. A predicate for which there is always some degree of change (no matter how small) which will affect whether or not this predicate is true of some particular case is an *intolerant* predicate. Say that a predicate which draws a sharp boundary between complementary cases is intolerant in this respect.

The thesis of strong tolerance (**ST**) drives the paradox of the A-sorites (and B-sorites) given in §3.2. Take the case of the maturation of the butterfly. For simplicity we can let $\emptyset$ represent the parameter of time. A small degree of change in respect of $\emptyset$ does not affect the correctness of applying the predicates 'is a caterpillar' to the butterfly-to-be at any stage in its development. Large differences in respect of $\emptyset$ will affect the correctness of applying this predicate, and small differences are cumulative (large changes can be formed from a series of small changes). These three features jointly ensure that it is correct to both apply and withhold the predicate 'is a caterpillar' to the butterfly-to-be at any stage in its development. Though the notion of strong tolerance is a key concept in the vagueness debate it cannot form the basis of a characteristic sentence. In §3.2, we saw that we could not employ the characteristic sentences **SS1** and **SS2** (nor the induction step **A2** or the premise **B1** from the B-sorites) to define vagueness. We likewise cannot employ Wright's (strong) notion of tolerance as the basis for a characteristic sentence for **ST** likewise induces paradox (at least given classical logic).[29]

This predicament is puzzling since, as we have seen with **SS1** and **SS2**, principles like **ST** appear to be initially plausible. One response to this puzzle—the very puzzle of vagueness— is to oppose the thought that vague predicates are tolerant in Wright's sense. On an

---

[28] I bracket consideration of those predicates such as 'is such that 6 is few' which are vague (neither clearly true nor clearly false) for all cases along $\emptyset$.

[29] It was noted above that a **A2** does not give rise to paradox in the subvaluational system of Hyde (1997), a system in which *modus ponens* fails. Likewise, **ST** does not give rise to paradox in this system. If one thinks that a substantial conception of vagueness has merit insofar as it respects our naive intuitions concerning vagueness, this feature ought to count as a distinctive merit of Hyde's view.

epistemicist view of vagueness, for instance, vague predicates are represented to be intolerant. For Sorenson (1988), as for other epistemicists, these predicates have 'unlimited sensitivity': there is some degree of change in respect of Ø which does make a difference as to whether F correctly applies or not to a particular case—it is just that we are *irremediably* ignorant as to which particular Ø-difference effects the change. A thesis of strong *intolerance* (**SIN**) says that every vague predicate is intolerant in this respect. A weak thesis of intolerance (**WIN**), in contrast, says that vague predicates have unlimited sensitivity, but that our ignorance as to which changes in respect of Ø will make a difference as to whether F applies is *remediable* ignorance (see the discussion of *optimistic* conceptions of vagueness in §3.8). We need to know if there is a view which is neutral as to which of **SIN, WIN, ST** is correct.

Say that

> F is *epistemically* tolerant with respect to Ø if and only if any sufficiently small positive degree of change with respect to Ø is insufficient to make a *clear* difference to the correctness of applying F to a particular case.

A thesis of epistemic tolerance (**ET**) entails that every vague predicate is tolerant in this respect. Roughly, epistemically tolerant predicates do not draw known or clear boundaries across their range of signification. When we add the observation that large changes in respect of Ø will affect whether F applies or not then, loosely speaking, we can say that across Ø there is a difference without a clear distinction. That is just what is (or ought to be) meant by the thesis that vague terms draw blurred boundaries. Given this, vagueness *qua* no clear boundary and vagueness *qua* epistemic tolerance can be thought of as effectively equivalent. (Henceforth, I will generally employ the latter terminology.)

Any theory which sanctions **ST** will sanction **ET**: to draw no (sharp) boundary is to draw no clear (sharp) boundary. Likewise any theory which represents vague terms to draw sharp but unknown or unknowable boundaries will also sanction **ET**: to draw no knowable boundary is to draw no known or clear boundary. **SIN** and **WIN** entail **ET.** Importantly, **ET** does not entail either **SIN, WIN** or **ST**: from the fact that vague terms do not draw clear boundaries we cannot infer whether such terms are (weakly or strongly) intolerant or strongly tolerant. Nor does **ET** entail that any small change with respect to Ø will fail to make a clear difference to whether or not F applies to a particular case. There is a world of difference between saying that no small degree of change makes a clear difference as to whether x is F

and saying that small degrees of change make no difference as to whether x is known to be F. Some formalisation might help here.

Suppose we have open sentence 'x is hot' governing the dimension of comparison of temperature where x ranges over the real numbers in the interval $0°$ to $100°$. Let 'C' abbreviate the functor 'It is clearly the case that'. To say that no small positive degree of change in x makes a clear difference as to whether x is hot can be given as: $\sim(\exists x)\ C(x\ \text{is hot})$ & $C(x\pm c\ \text{is not hot})$, where c is some suitably small value. In contrast, to say that small positive degrees of change make no difference as to whether it is clearly the case that x is hot can be given as: $\sim(\exists x)\ C(x\ \text{is hot})$ & $\sim C(x\pm c\ \text{is hot})$. The difference relies on the fact that 'C' and negation do not commute one way: $\sim C(x\ \text{is}\ F)$ does not entail $C(x\ \text{is}\ \sim F)$. To say that small positive degrees of change make no difference as to whether it is clearly the case that x is hot is simply an instance of the problematic schema **SS2** from which we can run the *B*-sorites. **SS2** is paradox-inducing while **ET** is not.

Such observations suggest how we might offer two (classically equivalent) characteristic sentences which exploit the notion of epistemic tolerance. Let's also be completely explicit about this and build in all the provisos we have encountered hitherto (as well as some provisos we have not previously discussed): Our characteristic sentences are

(ET1) $(\forall\alpha)(\forall\beta)$ if $|v(\beta)-v(\alpha)| <c$ and it is clearly the case that S is true in α then it is not clearly the case that not-S is true in β

which is classically equivalent to:

(ET2) $\sim(\exists\alpha)\ (\exists\beta)$ such that $|v(\beta)-v(\alpha)| <c$ and it is clearly the case that S is true in α and clearly the case that not-S is true in β

where for both **ET1** and **ET2**:

(a)   whether or not S is true depends on the value v in actual or counterfactual cases
(b) c is some small positive number (rational or real)
(c)   small v-values are cumulative (a series of small v-values forms a large v-value)
(d)   all the relevant v-facts are known, i.e. $(\forall\alpha)\ v(\alpha)$ is known in α
(e)   the meaning of S is known
(f)   we restrict the range of α and β to 'normal' cases of judgement conditions
(g)   where c is large then $(\exists\alpha)\ (\exists\beta)$ if $|v(\beta)-v(\alpha)| >c$ then S is clearly true in α and not-S is clearly true in β.

A sentence S is vague just in case when substituted in to either **ET1** and **ET2** these schemas are true—at least when all of the clauses **a-g** are satisfied. This may look rather cumbersome, but without each of these clauses we have not sufficiently isolated the constitution of vagueness. Let me discuss **ET1** and **ET2** first, before discussing each of these provisos in turn (where our discussion will bear a little repetition of the discussion given above)

It will be noted that the schemas **ET1** and **ET2** are effectively derived from the minimal margin for error principle **MME** given in Chapter One.[30] Just like the minimal margin for error principle **MME** one can motivate **ET1** and **ET2** on lightweight grounds. Across some smooth sorites series which is composed of cases of F at one end and cases of not-F at the other, one can reflect on the structure of one's knowledge and ignorance across such a transition and conclude that there are no two adjacent (or close) members of this series such that one is clearly F and the other is clearly not-F. However, there are also some key differences between **MME** and **ET1** and **ET2**.

Firstly, these schemas exploit a notion of clarity rather than knowledge. For our present purposes nothing turns on this: as mentioned above we can follow Williamson (1994, Ch.1) and take 'It is known that that' and 'it is clearly the case that' to be interchangeable. Secondly, these schemas concern whether or not a sentence S is clearly true in close counterfactual cases, not whether a condition C obtains. Again, nothing turns on this. One could equally formulate **ET1** and **ET2** using a variable for such a C condition and then speak of the linguistic vagueness of this sentence which features as a substitution for this variable (given a substitutional reading of the schemas). Thirdly, while **MME** as formulated in Chapter One, is general enough to cover cases of ignorance due to vagueness, the examples used there involved cases of inexact knowledge where vagueness was (implicitly) assumed not to be a factor. In standard cases of inexact knowledge, our ignorance is remediable when all the relevant v-facts are known. When one knows the exact height of the tree, one knows whether or not the tree is at least x feet high, for instance. It is for this very reason that certain of the provisos **a-g** are required.

Clause **a** we have encountered already. A more sophisticated formulation of **ET1** and **ET2** would advert to the fact that whether or not a vague sentence is true may depend on the variation in more than one underlying graded or continuous parameters. This is compatible with the vagueness of S issuing exclusively from the subject term or terms contained in S.

---

[30] Since Williamson's version of the epistemic theory of vagueness employs his margin for error principle **ME** to account for our ignorance in borderline cases, and since **ME** entails **MME** then Williamson's epistemicism entails the schema **ET1** (and **ET2**) in the case of vagueness

Clause **b** allows the underlying v-facts to vary continuously or discretely, while clause **c** is surely unproblematic.

Clauses **d-f** are crucial: they each ensure that our unclarity (or ignorance) does not result from the wrong source but solely from the vagueness of the sentence S. Clause **d** ensures that all the relevant v-facts are known. In the case of assessing whether the sentence 'My bath is hot' is vague, it is an prerequisite that I know the temperature of my bath (in actual and counterfactual cases). Likewise, nor must unclarity issue from any misunderstanding or ignorance as to the meaning of the sentence S. Clause **e** ensures just that. (What is meant by 'meaning' here is a matter which we will touch on in §3.8.) Clause **f** is also vital for one may know all the relevant v-facts and know the meaning of S and yet unclarity may result from reasons other than vagueness. For example, I may be drunk tired, distracted, or hallucinating such that **ET1** or **ET2** is satisfied. Indeed, external conditions may produce unclarity—the lighting might be bad, there might be smoke in the room, and so on. It may be a delicate matter to give a general characterisation of normal judgement conditions, and I will not endeavour to do so here beyond the remarks already given.[31] Lastly, clause **g** ensures that large differences in the value taken by v in α and the value taken by v in β will always entail that S and not-S are neither both true nor both false.

Say that a substituted sentence is vague just in case **ET1** (and **ET2,** together with the clauses **a-g**) are true for that substitution. Any conception of vagueness which does or can

---

[31] Williamson (1994, pp. 180-4), has expressed a doubt that one can specify normal (or standard) judgement or observation conditions in some substantive and non-circular manner. This challenge is problematic. The minimal conception of vagueness (and so every conception of vagueness) is also beholden to specify what normal viewing conditions must obtain in order to correctly isolate the phenomenon of vagueness. In claiming that it is constitutive of vague expressions that they draw no apparent sharp boundaries we must add the proviso (clause **f**) that this *non-appearance*, so to speak, is not due to bad lighting, boredom, inattentiveness, hallucination, dreaming, or misunderstanding on behalf of the phenomenological subject. Recall above, that in ostensively defining borderline cases in terms of unclarity, Williamson himself added the proviso that the unclarity in borderline cases stems from features of the statement involved. This is elliptical. It adds up to saying that our unclarity is *not* due to the non-optimality of the viewing conditions, or to abnormal cognitive functioning, or to any ignorance of the relevant v-facts, or indeed to any misunderstanding on behalf of the subject as to what the relevant term means. Without some idealisation of the epistemic or evidential conditions, it is not possible to strictly differentiate vagueness from other potential sources of *non-appearance* (or unclarity). Is this why, after all, Williamson stops short of (initially) offering anything more than an ostensive definition of vagueness? One might think that Williamson's caution is justified; if so, the promise of the minimal theory to furnish a generally acceptable minimal constitutive definition will be difficult to fulfil. But can this be right? It ought to be the main aim of any substantive conception of vagueness to explain why we are ignorant in borderline cases—why we do not know of certain adjacent members in some smooth sorites series whether or not they are F. If this ignorance were (partly) due to the evidential conditions being less than ideal, or to some cognitive malfunction on behalf of the subject, and so forth, then not only will we be unable to (constitutively) define vagueness, we will also be unable to satisfy the primary goal of any substantive account of vagueness. Specification of the normal judgement conditions (though an unforgiving and largely intractable business) is essential for this goal for any theory of vagueness. (See Hardin (1990) for some difficulties attending how to fix 'standard' viewing conditions for the particular case of colours. See Wright (1992a, pp. 12-3, pp. 120-4) on the some of the local difficulties in specifying the c-conditions for the euthyphronic contrast.)

constitutively define vagueness via **ET1** (or **ET2**) together with clauses **a-g** we may call a conception of vagueness *qua* epistemic tolerance. But can **ET1** and **ET2** avoid the problem of 'oldster' and cognate problems?

For simplicity I shall focus on **ET2**. As stated, **ET2** is in fact insufficiently general to ensure that the artificial term 'oldster' fails to count as vague. Why? Take the open sentence 'a person of age x is an oldster'. Given **ET2**, it is sufficient for this sentence to be vague that there is no x such that it is clearly the case that this sentence is true *and* clearly the case that the sentence 'a person of age x–c is *not* an oldster' is likewise true (given all the other provisos, and where c takes some small value). When the values of both x and x–c fall in the stipulated penumbral area (i.e. when $65 \leq x \leq 68$ and $65 \leq x-c \leq 68$) then both sentences are neither determinately true nor determinately false (according to the dictates of the stipulation), and so both are neither clearly true nor clearly false on just that basis (where we assume knowledge/clarity requires determinate truth). Hence, for those values the schema is satisfied. When the values of both x and x–c fall outside the stipulated penumbral area, then one (and only one) of the sentences is false, and so for these values the schema is satisfied. It is only when $x-c < 65 \leq x$, or when $x-c \leq 68 < x$ that we might—at first blush—expect these sentences to be clearly true.

Take the 'higher' of the two cut-offs. The statement 'a person of age x is an oldster' is clearly true, but the statement 'a person of age x–c is *not* a oldster' is not clearly true since it is penumbral—it is not determinately true (and not determinately false). Hence, the schema **ET2** is satisfied for the higher cut-off. Take the 'lower' cut-off. The statement 'a person of age x–c is *not* a oldster' is clearly true, but the statement 'a person of age x is an oldster' is not clearly true since it is penumbral. Hence, the schema is satisfied or the lower cut-off. Consequently, there are no two neighbouring values across the dimension of comparison of age for which the schema **ET2** fails: 'oldster' satisfies **ET2** together with **a-g**, and thus counts as vague. But we have seen that 'x is an oldster' is intuitively not vague: it draws a perfectly sharp and clearly identifiable three-fold division across its range of significance.

To meet this worry one might simply seek to offer the following schemas (which exploit the non-epistemic notion of determinacy used in the stipulation for 'oldster'):

(**ET3**) $(\forall\alpha)(\forall\beta)$ if $| v(\beta)-v(\alpha)| < c$ and it is clearly the case that S is determinately true in α then it is not clearly case that not-S is determinately true in β

which is classically equivalent to:

(ET4) ~ ($\exists\alpha$) ($\exists\beta$) such that $| v(\beta)-v(\alpha)| < c$ and it is clearly the case that S is determinately true in α and it is clearly the case that not-S is determinately true in β

(where all the other clauses **a-g** must be met in order for S to count as vague). The characteristic sentence **ET4** (and **ET3**) correctly identifies that the sentence 'a person of age x is an oldster' to be non-vague. In more detail, when $x-c \le 68 < x$ then 'a person of age $x-c$ is an oldster' is *not* determinately true, and it is, moreover, clearly the case that this is so, while the statement 'a person of age x is an oldster' is determinately true, and it is, moreover, clearly the case that this is so. Thus, for these values **ET4** is not satisfied.

But while **ET3** and **ET4** correctly predict that 'oldster' is non-vague, they are not able to identify the non-vagueness of any term which is stipulated to admit of second-order borderline cases. To illustrate: suppose we offer the following stipulation for 'oldster*':

(i) If x > 70 then 'x is an oldster*' is determinately determinately true,

(ii) If 70 ≥ x > 67 then 'x is an oldster*' is neither determinately determinately true nor determinately not determinately true,

(iii) If 67 ≥ x ≥ 66 then 'x is an oldster*' is determinately not determinately true and determinately not determinately not true,

(iv) If 66 > x ≥ 63 then 'x is an oldster*' is neither determinately determinately not true nor determinately not determinately not true,

(v) If 63 > x then 'x is an oldster*' is determinately determinately not true.

Again, there are no two neighbouring values on the dimension of comparison for which clause **ET4** (or **ET3**) fails. The point generalises. If we modify **ET3** and **ET4** to cope with this second-order artificial stipulation a third-order counter-example (with nine mutually exclusive and exhaustive truth-states) can be gerrymandered. In the limiting case, it is presumably possible to stipulate that some term is radically higher-order vague—that the borderline cases to borderline cases are non-terminating.[32] Is there a response?

The first thing to note is that counterexamples to **ET1-4** can only be generated within some suitable logical framework. For instance, the stipulation governing 'oldster*' requires a logic for the predicate 'determinately true' in which the following schemas are invalid:

---

[32] It's an interesting thought experiment to consider whether such a gerrymandered term would count as vague despite giving rise to a non-terminating hierarchy of borderline cases.

(DD)   If S is determinately true then S is determinately determinately true

(D~D)   If S is not determinately true then S is determinately not determinately true

(these are just formal mode analogues of the S4 and S5 principles, where 'determinately' takes the place of 'necessarily'). In a modal system which lacks these schemas (such as KT or KTB) there are no so-called reduction laws; hence there are an infinity of distinct modalities. Whatever logical framework is exploited in order to gerrymander some borderline term which nonetheless has sharp boundaries, we can define a variable τ which ranges over all possible modalities or 'truth-states' in this logical system. In KT and KTB, τ ranges over: {true, determinately true, not true, not determinately true, determinately determinately true, determinately not determinately true, ...}[33] In systems which contain reduction laws, τ will range over a finite number of modalities.[34] Given this, we can offer the following replacements for **ET3, ET4**:

(ET5) $(\forall\tau)(\forall\alpha)(\forall\beta)$ if $|v(\beta)-v(\alpha)| < c$ and it is clearly the case that S is τ in α then it is not clearly the case that S is not-τ in β

(ET6) $(\forall\tau) \sim (\exists\alpha)(\exists\beta)$ such that $|v(\beta)-v(\alpha)| < c$ and it is clearly the case that S is τ in α and it is clearly the case that S is not-τ in β

Schema **ET6** effectively says that there are no close cases in which it is clear that a sentence takes a certain truth-status in one case and clear that this sentence takes the complementary truth-status in another case. Schema **ET5** effectively says that if it is clearly the case that a sentence takes a certain truth-status then in nearby cases it is not clear that this sentence lacks this truth-status. (Given classical logic, both schemas are inter-derivable.)

The idea is that however far up the hierarchy of truth-states one goes in order to gerrymander some counterexample to **ET5** and **ET6**, one's stipulation will always invoke a sharp and clearly identifiable cut-off between at least one truth state τ and its complementary truth-state not-τ. The point also holds for any term which might be stipulated to be radically higher-order vague in the relevant sense. Thus such sentences as 'x is an oldster', 'x is an oldster*', and higher-order analogues, are correctly identified as non-vague. We have, at last,

---

[33] Note that the minimal theory does not entail that there are such (non-reducible) truth-states. The idea here is to combat the possibility that if there are such states then vagueness cannot be constitutively defined.

88

distinguished vagueness from superficially similar phenomena such as semantic incompleteness or underspecificity.[35] Sentences which satisfy **ET5** and **ET6**, together with clauses **a-g**, are vague, and conversely. Vagueness, from the perspective of the minimal theory is the phenomenon of epistemic tolerance: S is vague just in case there is no small change in respect of $\varnothing$ (actual or counterfactual) which makes a clear difference as to whether or not S is τ (where τ ranges over whatever truth-states S could possibly take over $\varnothing$).

We should thus be relatively satisfied that the characteristic sentence approach can after all offer a constitutive definition of vagueness which is acceptable to all parties to the dispute. Moreover we have succeeded in giving a definition of sentential vagueness which does not make an explicit reference to higher-order vagueness but which is rich enough to predict that terms (like 'oldster') which are stipulated to be have sharply defined borderline cases are non-vague. It looks as if there is no need to offer a substantive semantic story in order to constitutively define vagueness as Sainsbury (1991, p.174) has surmised.[36] The reluctance of Williamson, Sainsbury, and Tye (and many others) to offer anything more than an ostensive definition of vagueness now seems unjustified. What, then, are the conceptual and explanatory relationships between vagueness *qua* minimal tolerance and vagueness *qua* borderline cases?

3.7 Which dimension is more basic?

At the end of §3.2, it was suggested that vagueness *qua*-sorites susceptibility is secondary in the explanatory order—sentences are sorites-susceptible because they are vague, and not vice versa. At the end of §3.5, it was suggested that vagueness *qua* epistemic tolerance is more basic than vagueness *qua* borderline cases. Epistemic tolerance seems more basic than the epistemic notion of a borderline case (both explanatorily and conceptually): sentences give rise to borderline cases *because* they are epistemically tolerant and being epistemically

---

[34] In the modal system KT5 there are four reduction laws (see Chellas 1980, pp.147-154).

[35] Some commentators (e.g. Channell 1994, p. 2, *passim*) confuse vagueness *qua* tolerance with, what we may term, vagueness *qua* generality or 'underspecificity'. The predicate 'is between five and six hundred miles' is underspecific (in certain contexts it does not carry enough information), but it is not an example of vagueness proper in that this it draws clear boundaries over its range of signification. **10** allows us to distinguish between these distinct species of vagueness. (See Sorenson (1989) for more on the ambiguity of the term 'vague'.)

tolerant is constitutive of vagueness in a way that being a borderline case is not. It thus looks like we have already established how the three dimensions of vagueness are related to each other (both conceptually and explanatorily). But this is too quick. While it's fair to say that vagueness *qua* sorites-susceptibility is a less basic dimension of vagueness, in this section we will assess whether vagueness *qua* epistemic tolerance and vagueness *qua* borderline case are in fact conceptually equivalent dimensions despite the possibility of such terms as 'oldster'.

Generally speaking, most philosophers have been somewhat cavalier about what conceptual or explanatory relationships hold between the dimensions of vagueness *qua* borderline cases and vagueness *qua* tolerance. It's worth quoting in full what has been said about this matter. Black—though he in general tends to characterise vagueness in terms of the former dimension—is an early exception; he says:

> The finite area of the field of application of the word is a sign of its *generality*, while its vagueness is indicated by the finite [borderline] area and lack of specification of its boundary. It is *because* [my italics—P.G.] *small* variations in character are unimportant [...] that it is possible, by successive small variations, in any respect, ultimately to produce "borderline cases" (1937, p.31).

For Black then, vagueness *qua* tolerance is explanatorily (and presumably conceptually) prior to vagueness *qua* borderline cases.[37] Sainsbury, in speaking on behalf of what he calls the classical conception—the conception which 'characterises vagueness in terms of its allowing for borderline cases' (p.179)—also takes borderline cases to result from tolerance. He says:

> a very small difference in shade cannot make the difference between something being green and being blue, so we need a class of borderlines; a very small difference in age cannot make the difference between childhood and adulthood, so we need a class of borderlines (1991, p.168).[38]

---

[36] Indeed, Sainsbury in his (1990,1991) takes the hallmark of vagueness to be *boundarylessness*—a feature which surely entails that vague terms draw no *given* or clear boundary across their associated dimension and which ought therefore sanction the veracity of **ET5** and **ET6**.

[37] Black seems to have in mind something like Wright's strong notion of tolerance, when he should really have adverted to our notion of epistemic tolerance and said that it is because there is no small variation in character that is *known* to be important to the applicability of F that it is possible, by successive small variations, in any respect, ultimately to produce 'borderline cases' for F.

[38] In this paper Sainsbury aims to show that the classical conception generates an implausible model of higher-order vagueness. In his later paper of 1995b, as we have seen above, Sainsbury is happy to ostensively define vagueness via reference to borderline cases, though the focus of this paper lies elsewhere and there is no discussion of higher-order vagueness in this later paper.

While Hyde in discussing what he calls the 'paradigmatic conception'—the conception which aims to characterise vagueness in terms of borderline cases and borderline cases to those borderline cases—says:

> The paradigmatic concept we have been discussing initially attempts to accommodate the intuition that there is no apparent sharp boundary between the positive and negative extension of a predicate in terms of the presence of a penumbra or border region (or border cases). So, for example with the predicate "red" the absence of any apparent sharp boundary between the red and the non-red is initially described by reference to borderline cases (1994, p.36).[39]

In this case, Hyde seems to think that on this conception the notion of borderline case is more basic.[40] In contrast to Black and the conceptions outlined by Sainsbury and Hyde, one might assume that the two dimensions are conceptually equivalent and such that there is no proper explanatory priority to either dimension—we can equally well explain what vagueness amounts to by reference to either facet. This seems to be the view of Keefe and Smith:

> Clearly having fuzzy boundaries is closely related to having borderline cases. It might be argued, for example, that for there to be no sharp boundary between the F's and the not-F's just *is* for there to be a region of borderline cases of F. Our "two features" would then be thought of as the same central feature of vague predicates seen from two different slants (Keefe and Smith 1996, p. 3, fn.3, see also Keefe 2000, p. 7).[41]

Consider, then, the following two-part claim:

($\mathbf{ET} \Rightarrow \mathbf{CL}$) *From epistemic tolerance to borderline cases:* if a term is epistemically tolerant then this will entail that it will fail to draw a clear/known boundary across its range of signification and so there will be cases such that it is unclear whether or not this term applies.

($\mathbf{CL} \Rightarrow \mathbf{ET}$) *From borderline cases to epistemic tolerance:* if it is not known whether or not a vague term applies to certain cases then this will entail that this term gives no clear instruction as to where a boundary is to be drawn between such cases, and in the absence of a clear boundary no small change in the world will make a clear difference as to whether or not the term applies.

---

[39] What Hyde means by the 'paradigmatic conception' in many ways coincides with what Sainsbury means by the 'classical conception'. Like Sainsbury (in his 1990 and 1991 papers at least), Hyde is intent on undermining the grip which this conception has had upon the vagueness debate.

[40] There are other passages in Hyde which suggest that his own view of the matter is that vagueness *qua* epistemic tolerance is prior.

[41] Keefe and Smith fall foul of the confusion between 'no sharp boundary' (tolerance) and 'fuzzy boundary' (which they ought to have read as the feature of epistemic tolerance). Just as with Black, it is easy to adjust their comments and employ an epistemic notion of tolerance and an epistemic notion of borderline case.

Both Sainsbury's classicist and Black endorse **ET** $\Rightarrow$ **CL**, while Hyde's advocate of the paradigmatic conception advocates **CL** $\Rightarrow$ **ET**, and Keefe and Smith (in the quote given at least) endorse both claims. How can we adjudicate?

On an intuitive level at least, **ET** $\Rightarrow$ **CL** seems absolutely right. It is the route from (epistemic) borderline cases to (epistemic) tolerance that is suspect. A term (and its complement) can fail to apply (and so fail to clearly apply) to certain cases and yet nonetheless fail to count as epistemically tolerant—that was surely the lesson of the term 'oldster'.[42] So it looks like vagueness *qua* epistemic tolerance is more basic. But is this right?

Firstly, let's vindicate the naive intuition that vagueness *qua* minimal tolerance entails vagueness *qua* borderline cases. Suppose some open sentence '*Fx*' is epistemically tolerant such that it satisfies the schema **ET6** and clauses **a-g**. Let '*Fx*' abbreviate the sentence 'a person of x years of age is old' where '*F*', let us say, ranges over the series of natural numbers from 0 to 120. The relevant epistemic tolerance of '*Fx*' is such that no small drop in age makes a clear difference to whether or not '*Fx*' is τ. Let '*C*' abbreviate the functor 'it is clearly the case that', and let τ range over all possible truth-states which the sentence 'a person of x years of age is old' can possibly take over the dimension of comparison. Given the epistemic tolerance of '*Fx*' then we have

1   (1) $\sim(\exists x)\, C(\text{'}Fx\text{' is } \tau)\ \&\ C(\text{'}Fx{-}1\text{' is not } \tau)$

Since F also satisfies clause **g**, it follows that there is some large change in respect of $\varnothing$ which will make a clear difference as to whether or not '*Fx*' is τ, which we may conveniently give as follows:

2   (2) $(\exists x)\, C(\text{'}Fx\text{' is } \tau)\ \&\ C(\text{'}Fx{-}10\text{' is not } \tau)$

And for the sake of *reductio*, let us suppose that for all x, either it is clearly the case that '*Fx*' is τ or clearly the case that '*Fx*' is not τ:

3   (3) $(\forall x)\, C(\text{'}Fx\text{' is } \tau)\ \text{v}\ C(\text{'}Fx\text{' is not } \tau)$

---

[42] Indeed Keefe and Smith (1996, p. 3, fn.3) go on to recognise the possibility of terms (like 'oldster') which admit of borderline cases but where the borderline is sharply bounded. On p.15 they add that 'merely having borderline cases is not sufficient for vagueness: rather, with a genuinely vague predicate. the sets of clearly positive, clearly negative and borderline cases will each be fuzzily bounded'.

92

We now need to assume that for an arbitrary x that:

4   (4) $C($'Fx' is τ) & $C($'$Fx-10$' is not τ)

(i.e. line 4 is the typical disjunct for line 2). Given &-E we can thus infer:

4   (5) $C($'Fx' is τ)
4   (6) $C($'$Fx-10$' is not τ)

Now let us assume for the sake of absurdity that

7   (7) $C($'$Fx-9$' is τ)

By &-I on lines 6 and 7 we infer:

4,7   (8) $C($'$Fx-9$' is τ) & $C($'$Fx-10$' is not τ)

Given ∃-I this yields:

4,7   (9)   (∃x) $C($'Fx' is τ) & $C($'$Fx-1$' is not τ)

which contradicts line 1 and so we thus infer:

1,4   (10)   ~$C($'$Fx-9$' is τ)

Now an instance of line 3 is:

3   (11) $C($'$Fx-9$' is τ) v $C($'$Fx-9$' is not τ)

and by disjunctive syllogism on lines 11 and 12 allows us to infer:

1,3,4   (12) $C($'$Fx-9$' is not τ)

If we then assume for the sake of absurdity that $C($'$Fx-8$' is τ) by the same pattern of inference we can derive that $C($'$Fx-8$' is not τ). If we repeat this pattern of inference ten times, then we can thus derive:

1,3,4   (13) $C($'Fx' is not τ)

and given the factivity of the functor 'it is clearly the case that' this entails

1,3,4   (14)   'Fx' is not τ

Given factivity, line 5 entails

4   (15)   '*F*x' is τ

Contradiction. So reject 3 to infer:

1,4   (16)   ~($\forall$x) C('*F*x' is τ) v C('*F*x' is not τ)

which given the quantifier equivalences entails:

1,4   (17)   ($\exists$x) ~(C('*F*x' is τ) v C('*F*x' is not τ))

Which via de Morgan gives:

1,4   (18)   ($\exists$x) ~C('*F*x' is τ) & ~C('*F*x' is not τ)

and by $\exists$-E on lines 2,4 and 18 this gives:

1,2   (19)   ($\exists$x) ~C('*F*x' is τ) & ~C('*F*x' is not τ)

Result: **ET** $\Rightarrow$ **CL**. The fact no small $\varnothing$-change makes a clear difference as to whether or not x is *F* together with the fact that some large $\varnothing$-change does make a clear difference as to whether or not x is *F*, shows that '*F*x' gives rise to borderline cases *at least given the resources of classical logic*. This qualification is important. In the proof above, we have used principles (negation-introduction, the de Morgan's laws, disjunctive syllogism, and so on.) which have all been brought into doubt on certain approaches to the sorites paradox (see Chapter Four for more detail).[43] However, recall that in §3.1 it was argued that the minimal theory of vagueness is entitled to exploit classical resources until such point as this generates tangible controversy. Would any theorist (classical or otherwise) seriously seek to doubt the entailment from epistemic tolerance to epistemic borderline cases? Just because one rejects certain classical principles in order to combat the sorites does not entail that those principles fail throughout one's theory of vagueness. It seems right to say that the proof above is available to all partisans.

Of perhaps more immediate interest is that fact that it looks like line 19 can be generalised so as to form the basis of characteristic sentence for vagueness *qua* borderline cases as follows:

(CL3)   ($\forall\tau$) ($\exists\alpha$) In α, it is not clearly the case that S is τ and it is not clearly the case that
         S is not τ

---

[43] It's perhaps worth mentioning at this point that the step from line 16 to 17 is intuitionistically invalid. This is significant. The intuitionist can say that a vague term *F* draws no known boundary across its $\varnothing$-dimension does without saying that there is an object *a* for which it is not known that *a* is *F* and not known that *a* is *not-F*.

This schema effectively says that for any truth-state τ that S may take over $\varnothing$, there is at least one case such that it is not clear whether or not S is τ.[44]

What is notable about **CL3** in contrast to our earlier formulation **CL2** is that there is no explicit reference to higher-order vagueness. In **CL2** we had to employ a notion of **clarity** which carried with it the commitment to higher-order vagueness. What's attractive about **CL3** and both **ET5** and **ET6** is that these schemas can rule out the troubles with terms like 'oldster' and 'oldster*' without having to give an explicit model of (or commitment to) higher-order vagueness. The key question now is: does vagueness *qua* borderline cases as codified in **CL3** entail vagueness *qua* epistemic tolerance?

Suppose that our sentence 'Fx' satisfies **CL3**; thus for a particular number m it follows that:

1   (1) $\sim C('Fm'$ is $\tau)$ & $\sim C('Fm'$ is not $\tau)$

and suppose for the sake of *reductio* that 'Fx' is intolerant in the relevant respect, that is:

2   (2) $(\exists x)\, C('Fx'$ is $\tau)$ & $C('Fx-1'$ is not $\tau)$

from line (1) by &-E we get:

1   (3) $\sim C('Fm'$ is $\tau)$
1   (4) $\sim C('Fm'$ is not $\tau)$

Now let us assume for arbitrary x that:

5   (5) $C('Fx'$ is $\tau)$ & $C('Fx-1'$ is not $\tau)$

(i.e. line 5 is the typical disjunct for line 2). Given &-E, this yields:

5   (6) $C('Fx'$ is $\tau)$
5   (7) $C('Fx-1'$ is not $\tau)$

We know a priori that:

         (8)   x=m or x>m or x<m.

---

[44] Note that where τ just *is* determinate truth, then **CL3** is not satisfied by the open sentence 'x is an oldster' for there is no x for which speaker does not know whether or not 'x is an oldster' is determinately true or not. Nonetheless, **CL3** is satisfied by the sentence 'x is an oldster*'. Hence the need for the move made above: we must ensure that τ ranges over all possible truth-states invoked by whatever artificial stipulation we encounter.

If we suppose that x=m then we can immediately derive a contradiction on lines 3 and 6; so instead just suppose that

9   (9) $x > m$

But it is plain that if it is clearly the case that y is not old then it is clearly the case that y−1 is not old, which is to say:

10   (10) $(\forall y)\, C(\text{`}Fy\text{' is not } \tau) \supset C(\text{`}Fy{-}1\text{' is not } \tau)$

So given lines 7,9 and 10 we can prove given successive applications of universal instantiation and modus ponens that:

5,9,10   (11) $C(\text{`m' is not } \tau)$

which contradicts line 4 so we have:

1,5,9,10 (12) $\bot$

So suppose that:

13   (13) $x < m$

But since we know that if it is clearly the case that y is old then it is clearly the case that y+1 is also old, and so we have:

14   (14) $(\forall y)\, C(\text{`}Fy\text{' is } \tau) \supset C(\text{`}Fy{+}1\text{' is } \tau)$

So given lines 6, 13, and 14 we can infer:

5,13,14   (15) $C(\text{`}Fm\text{' is } \tau)$

which contradicts line 3 and so we have:

1,5,13,14 (16) $\bot$

and so by v-E on 8,9,12,13,16 we have:

1,5,10,14 (17) $\bot$

and by ∃-E on 2,5,17 we get:

1,2,10,14 (18) $\bot$

and by ~-introduction on 2, 18 we conclude:

1,10,14   (19)   ~(∃x) $C('Fx'$ is τ) & $C('Fx-1'$ is not τ)

Result: **CL** ⇒ **ET**. Thus we have shown that vagueness *qua* borderline cases (in the guise of **CL3**) entails vagueness *qua* minimal tolerance (on condition that the above rules of inference are valid in this context, of course). What conclusions can we draw?

It now looks like once we have properly isolated the phenomenon of vagueness *qua* borderline cases via the characteristic sentence **CL3**, then neither dimension is conceptually more basic: vagueness *qua* epistemic tolerance and vagueness *qua* borderline cases are indeed two facets of the same phenomenon. Our exposure to the problem of such terms as 'oldster' was in many ways a red-herring. Once we have isolated characteristic sentences which predicted the non-vagueness of such terms then any temptation to think that vagueness *qua* epistemic tolerance is the more basic phenomenon is lost. So while Fine (1975) may not have explicitly alluded to vagueness *qua* epistemic tolerance in setting forth his (indeterminist) account of vagueness *qua* borderline cases, it seems he (together with Williamson and Sainsbury) was nonetheless endeavouring to investigate the same phenomenon as Hyde, Burgess, Wright, and all those who have tended to focus on the phenomenon of blurred boundaries. Our partisans have at least from the outset been talking about the same thing all along.

3.8 Margin for error principles and realism

In §2.3, an argument was given which shows that at least insofar as one accepts that Williamson's margin for error principle **ME** is a necessary truth, then Williamson's model of inexact knowledge entails that there are unknowable truths. This was thought to be problematic for, given the supposition that the margin for error principle **ME** was to form part of the minimal theory of knowledge, it followed that epistemic minimalism appears to entail a realist conception of truth. While this is no doubt welcome to Williamson, it nonetheless seemed problematic for epistemic minimalism, since *prima facie*, at least, it would be quite surprising if our minimal theory of knowledge were able to resolve the realism/anti-realism debate in favour of the realist. Even though we then proceeded to reject Williamson's margin for error principle **ME** in favour of the weaker principle **MME** does this predicament remain for the weaker principle also? Consider the following argument:

97

(1)   It is known (via M) that: C obtains in α and C does not obtain in β

(where α and β are close). Given the distributivity of knowledge over conjunctions this entails:

(2)   It is known (via M) that C obtains in α and it is known (via M) that C does not obtain in β

and so by &-E we infer:

(3)   It is known (via M) that C obtains in α
(4)   It is known (via M) that C does not obtain in β

given the margin for error principle **MME**, from 3 we can infer

(5)   It is not known (via M) that C does not obtain in β

Contradiction on lines 4 and 5. So, reject 1 to infer:

(6)   It is not known (via M) that: C obtains in α and C does not obtain in β

Since, line 6 depends only on principles which are necessarily true, then by the rule of necessitation, and the modal equivalences, we can thus infer:

(7)   It is not *possible* to know (via M) that: C obtains in α and C does not obtain in β

Thus, just as with **ME**, given **MME** we can infer that there are undetectable truths. For example, by looking, it is not simply that I *do not* know at which exact time the tree ceased being less than ten feet high, I *cannot* know this fact (by looking). Likewise where C is the vague condition *the tree is tall*, there are *v*-values such that the value *v* takes in α is known but for which it is impossible to know in α the truth-value of the sentence 'the tree is tall'. Thus, **MME** when applied to vague statements entails that the truths values of (extensionally) vague statements are *impossible* to know.

Sainsbury (1995a, p. 591) has said that Williamson's margin for error principles tell us why 'knowledge of the boundaries of vague concepts is impossible'—that our ignorance in borderline cases is 'irremediable'. The same would seem to hold for the minimal margin for error principle **MME** when applied to the case of vagueness. But this looks like a most unwelcome consequence for the *quandary* view of vagueness given by Wright (2001). On the

quandary view the truth-values of borderline statements are unknown; but if true then they are in principle knowably so, and if false they are, likewise, in principle knowably so.[45]

Furthermore, it looks like the minimal theory of knowledge/vagueness rules out what may be termed an *optimistic* conception of vagueness under which a vague concept draws sharp but in principle knowable boundaries.[46] Who would dare sponsor such a view? Simons (1992, p. 168) comes pretty close when he says:

> [a]ccording to Williamson's theory [...] inevitably one party will be right, and the other wrong (though no one might be able to tell which). I agree that it can happen that one party is right and the other wrong, but that *such* mistakes can be discovered and rectified [...].

But our minimal theory of vagueness entails that, if there are such mistakes, it is impossible to discover them and so impossible to rectify them. Should we allow the minimal theory of vagueness to rule out quandarism and optimism from the outset? Indeed, should we let our minimal theory of knowledge entail a realist conception of truth? On the other hand, should we not simply reject the minimal margin for principle **MME** and the idea that one can define vagueness via the notion of epistemic tolerance?

In fact the dilemma just presented is a false one. Quandarism and optimism are compatible with the minimal theory of vagueness, and indeed the minimal theory of knowledge does not entail a realist conception of truth contrary to appearances. To show this, I'm merely going to concentrate on the particular case of vagueness. Let's look first at how the devotees of the epistemicist view of vagueness represent their own position and their own diagnosis as to why this position has been rejected or neglected.

The so-called epistemic view of vagueness has typically been taken to be committed to realism—but realism in just what sense? Cargile (1969, p. 200) takes it to be a characteristic feature of realism that 'in general, when a thing changes in time by losing or acquiring a property, it loses or acquires it instantaneously'. Campbell (1974) in contrast makes no particular reference to properties as such in his defence of the epistemic view; rather, epistemicism for him simply entails that there is, for example

---

[45] This does not entail that, for any vague statement S, either it is knowable that S is true or it is knowable that S is false, since one requires the principle bivalence (either S is true or S is false) to derive this principle from the schema: If S is true then it is knowable that S is true. It is a key element of the quandary view that one employs intuitionistic logic and semantics in order to be agnostic about the truth of the principle: either it is knowable that S is true or it is knowable that S is false.

[46] *Optimism* is also a position with the philosophy of mathematics, and the realism/anti-realism debate in general. See Shapiro (1997, pp. 203-11) and Tennant (1997, Ch.6).

a sharp cut-off point separating the heights of short men from those men not short [...] there will be a 'right' and a 'wrong' answer to the question of whether a borderline case is truly short man, despite the uncertainty that most competent speakers feel (1974, pp. 182-3).

Moreover, Campbell argues that to make sense of this view one must disavow verificationism and admit that there 'seems to be no inherent contradiction in the supposition that a proposition is true even though it is "in principle" impossible to discover whether it is true' (*ibid.*, p. 183). In a similar vein, Sorenson (1988, p. 45) has diagnosed the reasons behind the prevalent opposition to the epistemic view as issuing in part from 'a background of anti-realist theories of meaning'. Williamson (1994) is likewise clear that 'the epistemic view implies a form of realism, that even the truth about the boundaries of our concepts can outrun our capacity to know it' (*ibid.*, p. 4). For Williamson, the motivation behind any rejection of the (standard) epistemicist view stems from a belief in 'a suspect connection between what is true and what we can verify' (*ibid.*, p. xi). On all of these accounts, then, epistemicism is taken to imply a form of realism.

Dummett has identified the following principle, (which he has labels K, but I shall label **EC**), as essential to what he calls 'semantic anti-realism'[47]

(**EC**) If *S* is true, then it must be in principle possible to know that *S* is true

To reject **EC** is to endorse a conception of truth which is not epistemically constrained. However, as Dummett notes, one must be clear what is meant by 'in principle possible'. Williamson (1994, p. 212) in fact concedes that it is *metaphysically* possible to know the boundaries of our concepts in the sense 'that a being with cognitive powers greater than any we can imagine could know of someone with exact measurements *m* whether he is thin', but dismisses this possibility as irrelevant to the sort of conceptual knowledge possessed by ordinary speakers. Anti-realists likewise eschew any reference to ideal cognitive powers or ideal informational states in cashing out what is meant by 'knowability in principle'. But equally, such anti-realists urge that to be in principle knowable does not necessarily depend on what can *now* be known: they grant that there are statements which we do not at present

---

[47] Dummett (1976, reprinted 1993, p. 61). Stirton (1997), following McDowell (1976, p. 48), calls this the 'main thesis of anti-realism'. I take *semantic* anti-realism to be the view that there are global meaning-theoretical arguments which justify a refusal to assert bivalence.

know and will never in fact be able to know. Rather, the anti-realist denies that there are true statements which we *could not* recognise to be so.[48]

Sorenson and Campbell furnish no detailed explanation as to why we are ignorant of the truth-values of borderline cases. Beyond a general distaste for anti-realist theories of meaning, they do not provide a reason for thinking that sharp cut-off points are in principle unknowable. Williamson meanwhile hopes to substantiate just why we are ignorant, and just why this ignorance is irremediable, through applying his model of inexact knowledge, and the margin for error principle **ME**, to the case of vagueness.

In the case of vagueness, Williamson cannot locate the source of ignorance in borderline cases in any ignorance of the underlying sharp facts. By hypothesis, such facts are known in borderline cases. So where does our ignorance issue from? Williamson's reply is ingenious. It does not issue from any ignorance as to the meaning of the vague sentence in hand. Understanding an expression is not to be modelled along Fregean lines such that 'to grasp a sense is to know where its boundary runs' in conceptual space (1994, p. 210). Rather:

> The measure of full understanding is not possession of a complete set of metaphysically necessary truths but complete induction into a practice [...] to know what a word means is to be completely inducted into a practice that does in fact determine a meaning (*ibid.*, p. 211).

So where does our ignorance arise from if it does not arise from partial understanding? Well, the object of our ignorance are necessary truths of the following form: 'Every tree of height greater than m is tall', 'Every tree of height less than n is short', and so on. Crucially, the extension of 'tall', as determined by the speech community could easily have been slightly different, because the overall pattern of use of the speech community could easily have been slightly different. Extensions are sharp but unstable on this view. Since, it is an easy possibility that 'tall' could have determined a different extension, then even if one asserts something true when asserting 'Every tree of height greater than m is tall', it does not follow that one knows this necessary truth. Why? Because one's belief/or assertion is not safe from error—it could all too easily have been mistaken. Our knowledge of the extension of 'tall' is inexact—it is governed by a margin for error principle **ME**. That is why we lack knowledge of such necessary truths. While it is easy for the practice to determine a different extension, it is not an easy possibility for one's own use to get out of line with the use of the community. Hence, one can still know what 'tall' means (one's use aligns itself easily with that of others)

---

[48] See e.g. Dummett (1993, p. 61, 1973, p. 465).

without knowing some truth of the form 'Every tree of height greater than m is tall'.[49] Furthermore, as Sainsbury says above, our ignorance of the boundaries drawn by vague concepts is irremediable—it is *not possible* to know the extension of 'tall'—this is a direct consequence of the necessity of **ME**.

Rather than assess the merits of this intriguing proposal, we merely need to note that this account is not available on the minimal theory of vagueness. On the minimal conception of vagueness, there is no implication that vague terms draw sharp boundaries—such a conception should remain neutral as to the source of our ignorance in borderline cases. This effectively means that the *explanatory* basis for accepting **MME**, discussed at the end of the last chapter is not available. Our knowledge is not inexact *because* our method of acquiring knowledge in borderline cases are less than perfectly discriminatory. That would be to offer an epistemic explanation of ignorance, one which rules out the explanation which asserts that there is no fact of the matter to be known in borderline cases. Thus, only the *descriptive* route to **MME** is available in cases of vagueness. Even so, our minimal theory of vagueness still looks to be committed to the following conditional:

S is bivalent $\rightarrow$ $\Box\sim(\exists\alpha)(\exists\beta)$ [$K(S$ is true), via method M in α and $K(S$ is false), via M in β]

(where β is close to α, where 'K' abbreviates it is known that, where 'S' is the name of a vague sentence, and where 'M' is a rigid name for the method of belief formation employed). This is bad enough for both the quandary view and any optimistic conception of vagueness— this conditional, says that if S determines a bivalent boundary then this boundary is unknowable. Is there are possible response?

One immediate thought is that the sense of unknowability at play here pertains to *immediate* limitations on a speakers cognitive capacities relative, but does not pertain to what this speaker could know given unlimited investigative outlay and some indefinite and appropriate extension of their recognitional abilities. Notice that the modal status of both **MME** and **ME** is only assured once we have rigidified the name for the method of belief formation involved. Hence, given a fixed set of cognitive capacities, or in effect, a method or process of belief formation which is indexed to just those capacities, it ought to come as no surprise at all that certain truths are in principle unknowable. Their unknowability is anodyne.

---

[49] In this context, I bracket considerations of lucky knowledge and causal connections—relevant as these considerations are. Note that one cannot read off the necessary truth from the use, for Williamson thinks that

Take conceptual sources of inexactness. Given my present methods, I may not be able to read-off a principle of the form 'Every tree greater than m feet high is tall' from the use of the term 'tall' in the speech community. But that does not entail, that there is no method via which a speaker could come to have such knowledge—perhaps such a method will be discovered. The minimal theory of vagueness does not entail the thesis that the truth-values of borderline statements are knowable via some method or other, rather it is simply compatible with such an optimistic thesis. What it does entail is that it is not the case that there is some method via which the truth-values of borderline statements are knowable.[50] Hence the principles **MME** and **ME** do not support a realist conception of vagueness in any way. Furthermore, an optimistic conception of vagueness can exploit the margin for error principles **MME** and **ME** to explain why we lack knowledge of the truth-values of borderline statements. Likewise, Wright's quandary view is compatible with the minimal theory of vagueness, in which **MME** is taken to be a necessary truth, as we should expect.

3.9 Is there higher-order vagueness?

Russell (1923) and Black (1939) can be credited with putting the topic of vagueness back on the philosophical agenda. But notably these two philosophers disagreed about whether or not there is higher-order vagueness (*qua* borderline cases). While Russell says that the 'penumbra itself is not accurately definable' (1923, p.86), Black says it is 'impossible to accept Russell's suggestion that the fringe itself is ill-defined' (1939, p.37).[51] Though, most commentators have sided with Russell (though typically from within different conceptions of vagueness), some have sided with Black (though not always for quite the same reasons). Wright (1987, 1992b), for instance, has posed the challenge that independently of any worries concerning the sorites paradox, the notion of higher-order vagueness is in itself paradoxical. Kamp (1981) and Sainsbury (1990, 1991) meanwhile have raised doubts concerning any model of higher-order vagueness couched in set-theoretical terms. Koons (1994), in the context of defending a particular hybrid conception of vagueness, has said that the

---

'meaning may supervene on use in an unsurveyably chaotic way' (1994, p. 209). Indeed, there is no requirement that your use of 'tall' align perfectly with mine in order to have full mastery of the term.

[50] It is a quantifier shift-fallacy to reason from 'every truth is knowable via some method or other' to 'there is some method via which every truth is knowable'. To deny the latter claim is thus not to deny the former.

[51] Black's doubts stem from the thought that classical negation rules out the possibility of an indeterminist conception of borderline cases. These doubts relate to those raised by Williamson (1992, 1994, Ch.7) concerning the stability of truth-value gaps (see Chapter Four for more discussion).

desire to avoid sharp boundaries is no reason to postulate higher-order vagueness. I will not postulate such second- or higher-order vagueness unless some independent argument can be made for doing so (Koons 1994, p. 447).[52]

Burgess (1990, 1998), in contrast, grants the existence of lower-order vagueness (nth-order vagueness for small n-values) but has questioned the existence of nth-order vagueness for all n. For Burgess,

> it is far too early in the day to claim with confidence that higher-order vagueness fails to terminate, either as a matter of logic or as a matter of fact' (1998, p. 240).

And indeed, like Koons, he defends this view from within a hybrid conception of vagueness, whereby

> for each [vague] concept, at some point in the ascending orders of vagueness, higher-order vagueness will terminate for it [...] Ordinary speakers could not know where this order is, still less could they know the exact location of these lines' (*ibid.*, pp. 249-50).

For Burgess, vague expressions give rise to non-epistemic indeterminacy in the guise of borderline cases, but this indeterminacy is fairly shallow, as it were, since it does not generate a non-terminating hierarchy of borderlines cases, but rather a partial hierarchy which terminates at some unknowable point. Thus, at some level in the hierarchy the borderline cases to the borderline cases will have sharp but unknowable boundaries.[53]

In this last section, we will thus be concerned with two challenges: that there is first-order vagueness but no nth-order vagueness for $n>2$ (Koons and Wright), and that there is nth-order vagueness for small n-values (what we may loosely term *lower-order* vagueness) but no radical higher-order vagueness—nth-order vagueness for all n (Burgess).

Consider then the following principle:

(**CC**)   ($\forall\alpha$) If $C^{n-1}p$ in α then $C^n p$ in α (for $n>1$)

---

[52] Koons thinks that there is vagueness *qua* non-epistemic borderline cases at first-order such that vagueness gives rise to truth-value gaps. However, he maintains that the limits of the gap are unknowable. It may in the end be that Koons is merely reluctant to postulate *non-epistemic* but not epistemic higher-order vagueness.

[53] I hasten to add that Koons and Burgess sponsor quite different hybrid conceptions of vagueness. While there may be advantages to be had from adopting a hybrid view of vagueness, the most natural point for the orders of non-epistemic borderline cases to terminate is at first-order.

(where 'C^{n-1}' abbreviates $n-1$ iterations of 'It is clearly the case that' (for $n>1$), and where 'p' does not contain an occurrence of 'It is clearly the case that'). (Thus, where $n=2$, and 'p' abbreviates 'x is hot', for example, then we have: $(\forall\alpha)$ If it is clearly the case that x is hot in α then it is clearly the case that it is clearly the case that x is hot in α.) If this principle fails for $n=2$, then there must at least be second-order vagueness. If this principle fails for all n, then there must be *radical* higher-order vagueness, i.e. a non-terminating hierarchy of borderline cases. Burgess (1990, 1998) effectively argues that while there is lower-order vagueness (i.e. **CC** fails for small n values), the orders of vagueness terminate at some unknowable point (i.e. there is some largish value for n for which **C** is valid for all values $>n$.)

Hyde, in contrast, claims that radical higher-order vagueness *qua* borderline cases arises

> because vague predicates typically fail to draw *any* apparent sharp boundaries within their range of signification (Hyde 1994, p. 36).

Is he right to do so? Arguably, yes. If there is epistemic tolerance at each order n (which is just shorthand for saying that such sentences as '$C^{n}p$' are epistemically tolerant for any n-value) then **CC** will fail whatever value we take for n. Loosely, if epistemic tolerance goes all the way then so should genuine higher-order vagueness *qua* borderline cases. To show this suppose we have some vague sentence S which says that p (where S is a sentence of the object-language and thus S does not contain the operator 'clearly'). No matter how many iterations of the C-operator we prefix to 'p' the result is always a sentence '$C^{n}p$' which is likewise epistemically tolerant. To show that this is genuine, non-terminating higher-order vagueness we must ensure that **CC** fails *for all n-values*, thus ensuring, contra Burgess (1990,1998), that genuine higher-order vagueness does not terminate at some sharp (and unknown) point.

To do this we can simply reconfigure Wright's paradox of higher-order vagueness in epistemic terms (Wright's original paradox is given in terms of a non-epistemic 'definitely' operator.) Once we do that we can see that it is no paradox at all, but rather it merely shows that the 'clearly' operator is non-iterative, i.e. **CC** fails *for all n values*.

We can certainly say that if epistemic tolerance goes all the way up then the following schema ought to hold (for any $n\geq1$)

(1) $C(\sim(\exists\alpha)\,(\exists\beta)\,CC^{n-1}p$ in α and $C\sim C^{n-1}p$ in $\beta)$ (where β is close to α)

Given the factivity of 'it is clearly the case that' then this entails:

(2) $\sim(\exists\alpha)(\exists\beta)\ CC^{n-1}p$ in α and $C\sim C^{n-1}p$ in β (where β is close to α)

Now let us take it as given that:

(3) $C\sim C^{n-1}p$ in β

and also assume for reductio that

(4) $C^{n-1}p$ in α (where β is close to α)

Given **CI**, i.e. the rule:

(**CI**) $\dfrac{\Gamma \vdash C^{n-1}p\ \text{in}\ \alpha}{\Gamma \vdash C^{n}p\ \text{in}\ \alpha}$ (where all the members of the premise set Γ are prefixed with 'C', and where $n>1$)

(note that **CI** is just the 'rule-of-inference form' of **CC**) then from line 4 we can infer:

(5) $CC^{n-1}p$ in α

and by &-I on 3 and 5 and two applications of $\exists$-I this yields:

(6) $(\exists\alpha)(\exists\beta)\ CC^{n-1}p$ in α and $C\sim C^{n-1}p$ in β (where β is close to α)

which contradicts line (2) and so we reject 4 to infer:

(7) $\sim C^{n-1}p$ in α

and by another application of **CI** we can infer:

(8) $C\sim C^{n-1}p$ in α

and by a step of conditional introduction together with two steps of $\forall$-I, this yields:

(9) $(\forall\alpha)(\forall\beta)\ C\sim C^{n-1}p$ in $\beta \to C\sim C^{n-1}p$ in α

which depends only on line 1. But 9 is disastrous for it allows us to infer that if $C\sim C^{n-1}p$ in some case α, then $C\sim C^{n-1}p$ is true in all cases. We have four basic options:

(a) Retain **CI** for all $n>1$ and thus say that '$C^{n}p$' is not epistemically tolerant for any $n>0$

(b) Retain **CI** for $n>m$ and thus say that '$C^{n}p$' is not epistemically tolerant for $n>m$

(c) Retain **CI** for all *n*>1 but reject one or more of the other rules of inference

(d) Reject **CI** for all *n* (and so, given the deduction theorem, reject **CC** for all *n*).

Let's briefly take each of these in turn. Option (a) is surely the least attractive. It ought to be entirely uncontroversial that the sentence 'x is clearly hot' is vague (on our three dimensions). Indeed, that is so even if one takes 'clearly' to be a redundant operator and thus, by default, a non-vague operator. Option (b) is perhaps one which Burgess might fall back on. It amounts to saying that not only does **CC/CI** hold for *n*-values greater than *m* (where *m* is, say, greater than 10) but that, as the above proof shows, this means that '$C^n p$' ceases to be epistemically tolerant at level *m*+1, though the exact value for *m* remains unknown. But this places the burden of proof on Burgess to produce a value for *n* for which it is known that '$C^n p$' is not epistemically tolerant. Equally, given the connection between epistemic tolerance and sorites-susceptibility, then Burgess must also argue that '$C^n p$' is not sorites-susceptible for some *n*-value. But we can always employ some sentence of the form $C^n p$ to generate a sorites paradox (if the sentence that says that *p* is itself sorites-susceptible). Consequently, (b) is no option either.

Option (c) is perhaps more plausible still. In endeavouring to show that there is genuine radical higher-order vagueness we have made use of rules of inference which have been brought into question when dealing with vagueness—*reductio ad absurdum*, conditional proof, being the two most obvious ones. So it's certainly true that one might seek to reject the import of the above proof by questioning the use of these rules from within some substantive non-classical theory of vagueness. But Black seems to accept classical logic, and Koons (1994) appears to sponsor classical logic in the meta-language. Even if one does think that the sorites paradox is to be defused from within some non-classical logic that does not entail that classical rules of inference should thereby fail in the context of showing that there must be higher-order vagueness. So at the very least the above derivation represents a challenge to find a 'relevant failing' in the rules of inference employed. This leaves us with the last option.

If **CI** fails for all *n* then **CC** fails for all *n* (given the deduction theorem), and as we have seen, that entails that there must be radical *n*th-order vagueness whatever value *n* takes. If that is right, then from axioms which are uncontroversial (roughly, vagueness is epistemic tolerance) we can derive an important and controversial theorem within our minimal theory of vagueness. The promises of our minimal theory have been satisfied: we can give a constitutive definition of vagueness which successfully distinguishes vagueness from various

distinct but related and phenomena, and on that basis we can show that (radical) higher-order vagueness is not an illusion as some have thought but a phenomenon that we are all beholden to recognise.

What are the main results of this chapter? We have found a way to give a relatively neutral and constitutive characterisation of vagueness in terms of the phenomenon of epistemic tolerance. Vagueness just *is* epistemic tolerance. Hence, we have found a way to distinguish the vague from the non-vague. Moreover, we have rigorously shown (at least given classical logic) that vagueness *qua* epistemic tolerance and vagueness *qua* epistemic borderline cases is just the same phenomenon. Lastly, we have seen that if we allow that no matter how many times we iterate the operator 'It is clearly the case that' (where these iterations are prefixed to some first-order vague sentence), if the result is an epistemically tolerant sentence, then there must be radical higher-order vagueness.

CHAPTER FOUR
TRUTH-MINIMALISM AND TRUTH-VALUE GAPS[1]

Chapter 4: Truth-minimalism and truth-value gaps

4.1 Truth-theoretical and proof-theoretical problems for gappy logics

4.2 Truth-minimalism and the transparency platitude

4.3 Truth and proof for truth-value gaps

4.4 Penumbral connections and Wright's challenge

Is minimalism about truth compatible with truth-value gaps? Is it even possible to have a logic and semantics for truth-value gaps which is both truth-theoretically and proof-theoretically well-motivated? Can truth-value gaps feature in a substantial theory of what it is to be a borderline case? These are the three main questions dealt with in this chapter.

In §4.1, I set forth a range of arguments which allegedly reveal the truth-theoretic and proof-theoretic weaknesses of any 'gappy' logic. In §4.2, it is argued that truth-minimalism and truths-value gaps are perfectly compatible. One key thought here is that the so-called transparency property of truth does not hold in full generality (as Wright and others have argued). In Chapter One, I drew a distinction between deflationary minimalism (truth cannot feature in explanation) and inflationary minimalism (truth can feature in explanations). If the arguments in this section are correct, it looks like we should favour the latter form of minimalism. In §4.3, I develop a logic and semantics for truth-value gaps which is both truth-theoretically and proof-theoretically well-motivated (despite the many doubts raised by Williamson, Machina, Horwich, and others). One key move in this section is to express the T-schema using a non-contraposible extensional conditional allowing the deduction theorem to be retained in full generality. Though most of the given proof-theory is available to the supervaluationist, the rule of disjunction-elimination is taken to be valid and so the semantics for disjunction is truth-functional. (With respect to vagueness, the penumbral connections are thus not-validated.) A novel three-valued truth-functional logic results, one which is arguably more satisfactory than any competitor three-valued logic.

---

[1] This chapter is based on a talk given at the *Logica '99* conference, Liblice, Czech Republic, June 1999. (This talk was later published as 'Anti-realism and the liar paradox', *The Logica Yearbook 1999*, (ed) T. Childers, Prague: Filosofia.) Many thanks to Göran Sundholm, Stewart Shapiro, and Alan Weir for their constructive comments on that occasion. Patrice Philie, Stephen Read, and Sven Rosenkranz provided invaluable comments on an earlier draft.

In §4.4, I offer some reasons why the penumbral connections are not validated. It is argued that our intuitions as to whether or not the penumbral connections should be sanctioned by any respectable theory of vagueness are in any case hampered by an important, though entirely neglected, challenge given by Wright (1995). The nub of this challenge is that indeterminacy in truth-value (*qua* borderline case vagueness) ought to be a status compatible with the poles of truth and falsity—a challenge that, if correct, would rule out any three-valued, many-valued, or supervaluational conception of vagueness from the start. In reply to this challenge, it is argued that one can both respect Wright's challenge *and* find a place for truth-value gaps in a theory of vagueness, should one wish to do so.

4.1 Truth-theoretical and proof-theoretical problems for gappy logics

Perhaps the most sophisticated and sustained critique of any logical system in which truth-value gaps are admitted has been given by Williamson (1994, pp. 187-92). Williamson's basic thesis is simple: to deny bivalence is inconsistent in the presence of Tarski's T-schema for truth, and yet to reject this schema is to have no theory of truth at all. A simplified version of Williamson's argument runs as follows: suppose we have some sentence S which says that p. To assert that S is neither true nor false is to assert

(1)   not: either S is true or S is false

As this sentence says that p, then it is subject to the following familiar disquotational principles:

(T) S is true if and only if p
(F) S is false if and only if not-p

We can then substitute the right-hand sides for the left-hand sides of **T** and **F** in (1), yielding:

(2)   not: either p or not-p

and by the relevant de Morgan's law we derive the contradiction

(3)   not p and not not p

A contradiction is also derivable in the material mode of speech by assuming that it is neither true nor false that *p* in the presence of the following 'equivalence schemas':

(EST)   It is true that *p* if and only if *p*

(ESF)   It is false that *p* if and only if not-*p*

The simplicity of Williamson's argument is seductive. Indeed, Burgess (1998, p. 248) has claimed that 'nobody disputes the formal details of Williamson's derivation'. Moreover, this argument is given a very high profile in Williamson's case for the epistemic view of vagueness.[2] Those who have sought to reject the epistemic view (in its classical form) have acknowledged the urgent need to block the argument (e.g. Travis 1999, fn1).

The immediate response to argument such as Williamson's is to hold that the T-schema is in any case to be rejected if one is to admit statements which lack a truth-value. It is well known that Dummett has argued (in many places) that the T-schema is incompatible with truth-value gaps. Here's one example of Dummett's point:

> It is necessary to admit counter-examples to the schema (**T**) in any case in which we wish to hold that there exist sentences which are neither true nor false: for if we replace [the right-hand side] by such a sentence, the left-hand side of the biconditional becomes false [...], although, by hypothesis, the right-hand side is not false (Dummett, 1978, p. 233).[3]

The same, arguably, goes for **F**: if the right-hand side of **F** is gappy, then the left-hand side is false, so left-hand side and right-hand side do not match in truth-value. For Williamson, we cannot use this observation to bolster the case for truth-value gaps since 'it does nothing to meet the rationale for (**T**) and (**F**)' (1994, p.190). What is that rationale? It derives from a well-known remark by Aristotle, which runs: 'To say of what is that it is not, or of what is not, that it is, is false, while to say of what is that it is, or of what is not that it is not, is true'. Williamson glosses this as follows:

> Given that an utterance says that TW is thin, what it takes for it to be true is just for TW to be thin, and what it takes for it to be false is for TW not to be thin (1994, p. 190).

---

[2] Of course given the possibility of hybrid views of vagueness such as the view given by Koons (1994), which we briefly encountered in the last chapter, one might wonder why Williamson seeks to undermine the thesis of truth-value gaps. Williamson's argument is not new (though Williamson's discussion of just why this argument is sound is distinctive). It seems to appear in various guises in a number of places, but perhaps most explicitly in McCall (1970, p. 83).

There is, however, a stock response to the predicament faced by the devotee of truth-value gaps, which, on the face of it, appears to allow this rationale to be satisfied while admitting truth-value gaps. The idea is that even though left-hand side and right-hand side of the schemas **T** and **F**, and the schemas **EST** and **ESF**, do not bear a relationship of material equivalence they nonetheless bear a relationship of mutual entailment (van Fraassen 1966, p. 494) or 'inter-deducibility' (Smiley 1960, p. 129; McCall 1970, pp. 84-5; Keefe 2000, p. 214-7). Thus, we have the following surrogates (given in terms of deducibility) for **T** and **F**:

(**T1**)   Given $\ulcorner S$ is true$\urcorner$ infer $\ulcorner p \urcorner$ Truth-elimination
(**T2**)   Given $\ulcorner p \urcorner$ infer $\ulcorner S$ is true$\urcorner$ Truth-introduction
(**F1**)   Given $\ulcorner S$ is false$\urcorner$ infer $\ulcorner$not-$p\urcorner$ Falsity-elimination
(**F2**)   Given $\ulcorner$not-$p\urcorner$ infer $\ulcorner S$ is false$\urcorner$ Falsity-introduction

which are supposed to replace the material reading of 'if and only if' in **T** and **F**. (Likewise, analogue rules hold for the schemas **EST** and **ESF**.) On that basis, the thought goes, one can block the step from lines 1 to 2 in Williamson's proof, since while a material reading of **T** and **F** sanctions the substitution step, a mutual entailment/deducibility reading does not.

That's a well known response, but is it at all satisfactory? Williamson doesn't address this response in his discussion following the presentation of the above argument (given in Ch. 7 of his 1994). He does briefly address this response in an earlier Chapter on supervaluation. He says:

> the mutual entailment reading fails to capture the disquotational idea. If the truth predicate really does have the effect of stripping off quotation marks, then the material biconditional that 'A' is true if and only if A strips down to the tautology that A if and only if A. The supervaluationist denies that supertruth behaves like that; the availability of the mutual entailment reading is an irrelevance (1994, pp. 162-3).

These remarks are rather cryptic, but the general idea seems to be that the disquotational properties of truth ought to be entirely uncontroversial. But, in the quote just given we are given no reason for thinking that A and 'A is true' are inter-substitutable *salva veritate* in all extensional contexts—why should we think that it is trivially true that we can strip off quotes within the scope of negation for instance? The mutual entailment reading is indeed an irrelevance, but not for the reasons that Williamson advances. For all that van Fraassen, Smiley, McCall, and Keefe have said, they have not ruled out that the entailments **T1/2**, **F1/2**

---

[3] Dummett (1978, p.4) runs a similar argument against **EST**.

contrapose. Typically of course, entailments do contrapose, which is to say that when A ⊨ B then we can infer ~B ⊨ ~A (*cf.* Priest 1987, p.109; Edgington 1993, p.195). But then it looks as if when p ⊨ S is true then S is not true ⊨ not-p; likewise, when not-p ⊨ S is false then S is not false ⊨ not-not-p. But then the contradiction returns: if S is not true it follows that not-p, but if S is also not-false it follows that not-not-p. This is important, for it suggests that it is not the *material* reading of **T** and **F** that is driving Williamson's proof, but the fact that **T** and **F** are contraposible. If they are contraposible then the substitution step from lines 1 to 2 ought to be valid—whether one reads the if and only if in these schemas as a material relation or as relation of mutual entailment. Can one then express **T** and **F** as non-contraposible material conditionals in order to block the substitution step in Williamson's proof? That is a question to which we will return to in §4.3.

Let's pretend (for the sake of argument) that the mutual entailment reading furnishes the devotee of truth-value gaps with both an acceptable rejoinder to Williamson's proof and an acceptable theory of truth (one which meets the Aristotelian rational for **T** and **F**). To read **T** and **F** as mutual entailments (but not as material equivalences) means that the deduction theorem (and the corresponding rule of conditional proof) is invalid. Why? Because it is allowed that p ⊨ S is true but not allowed that ⊨ $p \rightarrow S$ is true. But surely the deduction theorem captures the absolutely basic connection between suppositional reasoning and the categorical assertion of conditional claims (cf. Williamson 1994, pp. 151-2; Mackie 1973, Ch. 5). Any logical system in which the deduction theorem is invalid can hardly be said to be a logic at all.

Furthermore, Machina (1976, pp. 52-3) has given an argument which appears to show that a denial of bivalence leads to a contradiction even if **T** is read as a mutual entailment. A simplified version of his argument can be given as follows: Suppose S lacks a truth-value; then:

(1)* S is not true
(2)* S is not false

and for the sake of *reductio* suppose

(3)* p

Given the rule of truth-introduction **T2**, we can infer

113

(4)* S is true

which contradicts (1)* and so by negation-introduction we can discharge (3)* to infer

(5)* not-p

and given the rule of falsity-introduction entailment **F2**, we can infer

(6)* S is false

which contradicts (2)*. Hence, the supposition of a counterexample to bivalence is self-inconsistent via apparently trivial logic. The immediate response is to reject the rule of negation-introduction (*reductio as absurdum*). But in so doing, we appear to have lost a basic and natural inference pattern which allows us to reject a premise of an argument if we can derive a contradiction from the premise-set. Indeed, if negation-introduction is not universally valid then certain derived rules of inference are likewise not universally valid. The two most obvious rules are contraposition and *modus tollens*. In itself, the devotee of truth-value gaps might parade this last fact as a virtue since it provides an answer to the worry, raised above, that if entailments contrapose then the substitution step in Williamson's proof ought to be valid even if we read **T** and **F** as expressing a relationship of mutual entailment. Here the thought is that once contraposition is rejected the mutual-entailment reading comes properly into force: given $p \vDash S$ is true, then it does not follows that S is not true $\vDash \sim p$. But this simply brings our earlier puzzle into greater relief: if at root the stability of truth-value gaps turns on whether or not the schemas **T** and **F** are contraposible, of what possible help is the mutual entailment reading *per se*? Surely the aim of the gappist should be to see if we can assert $\vDash p \to S$ is true without thereby being committed to assert $\vDash S$ is not true $\to$ not-p? The burden of proof on the gappist is clear: to show that we can indeed do just that. But the gappist must also do more than this: they must reconstruct a proof-theory for a gappy logic which respects the sort of inference patterns we find in natural language, as Williamson demands.

Supervaluational logic also suffers from local proof-theoretical problems. The rules of v-elimination (argument by cases) and the rule of ∃-elimination are also invalid if the famous

114

penumbral connections are to be respected.[4] If 'p v not-p' is true (even when 'p' lacks a truth-value), then given the rules of **T2** and **F2**, one can infer, using v-elimination, that 'S is true or S is false'. But bivalence is rejected in supervaluational semantics, and so v-elimination is not truth-preserving. Likewise, the supervaluational response to the A-sorites is to deny the induction step which entails the claim 'it is true that there is a case α and a case β such that a is F in α and a is not-F in β' (where β is close to α). Given the rule of $\exists$-elimination (together with the rules of truth-elimination and truth-introduction) one can then infer that 'there is a case α and a case β such it is true that: a is F in α and a is not-F in β' (where β is close to α). It is integral to the supervaluational response to the sorites to deny this latter claim since it commits a speaker to a determinate cut-off. Hence, $\exists$-elimination is not truth-preserving.[5]

Before attempting to meet these truth-theoretical and proof-theoretical worries, let us ask: is it worthwhile for the minimalist to proceed—for if minimalism and truth-value gaps are incompatible then from the perspective of minimalism, it is irrelevant to know whether there is a well-motivated gappy logic? So let us first establish the compatibilist thesis.

4.2 Truth-minimalism, truth-value gaps, and the transparency platitude

Are truth-value gaps available to the minimalist? The compatibilist says *yes*, while the incompatibilist says *no*. Horwich (1998, pp. 76-7) falls in the latter camp, while Holton (2000) falls in the former. Wright sketches how a minimalist might accommodate truth-value

---

[4] These connections are the logical relations—most notably exclusivity and exhaustivity—which hold among sentences which share a common borderline and yet which are indeterminate in truth-value. (Statements are indeterminate in value, on the standard supervaluational view, when they are neither true nor false.) Take the penumbral connection of exhaustivity. If an object a is on the red–orange borderline, then arguably there are only two candidates for the colour of a: red, orange. Thus, a is either red or orange, despite the fact that the sentences 'a is red' and 'a is orange' are indeterminate in truth-value. These sentences also bear the penumbral connection of exclusivity to each other: being red entails being not orange and being orange entails being not-red. Since red and orange are thus exclusive and exhaustive it follows that a is either orange or not-orange. Thus the fact that a is a borderline case does not invalidate the law of exclude middle. Where a is borderline for the predicate 'is small', however, then 'either a is small or red' is not a true disjunction since there is no penumbra, and *a fortiori* no possibility of a penumbral connection, between red objects and small objects. The argument is originally due to Dummett (1975), but is also employed by Fine (1975), and in a rather more rigorous form by Sanford (1976).

[5] The rules of v-elimination and $\exists$-elimination are also invalid in the gappy logic given by McCall (1970). If one takes the introduction and elimination rules to give implicit definitions of the logical constants then to lack such rules leaves these constants undefined. Since supervaluation cannot define the logical connectives via three-valued truth tables (since it is a non-truth functional logic) then it must find some other way of specifying the meaning of the logical connectives.

115

gaps in his (1992a, pp. 61-4), but appears far more sceptical about compatibilism at other places (1992a, p. 32, 2000, pp. 61-2). Field (1992, p. 332, fn.1, 2001, p. 222, fn.2) is also somewhat equivocal, though he seems to be a reluctant compatibilist (at least, Field thinks, if one wants the resources to combat the semantic paradoxes).

To adjudicate between these various positions, we must in any case recognise two different kinds of compatibilism/incompatibilism corresponding to two different sorts of truth-value gaps which in turn correspond to two different sorts of truth, namely, *weak* truth and *strong* truth (together with the correlate notions of falsity). The distinction between strong truth and weak truth is a familiar one (see e.g. Dummett 1978, pp. 4-5; Hugly and Sayward 1992). On the weak reading of truth/falsity, either side of the schemas **T** and **F** (and the schemas **EST** and **ESF**) co-vary in truth-status, while on the strong reading (the reading employed by Dummett in the quote given above) the left-hand side of these schemas is false when the right-hand side lacks a truth-value.

Holton has defended compatibilism with respect to the weak reading of truth. But, as Holton concedes, the weak reading precludes us from saying that S is neither true nor false for this is just to say that $\sim p$ and $\sim\sim p$, (given the validity of **T** and **F** on a weak reading). Moreover, it precludes us from asserting or denying that bivalence holds for all meaningful sentences (some of which may admit of truth-value gaps). But surely the whole point of recognising truth-value gaps is to employ the notion of gappiness for theoretical work in semantics—to explain such phenomena as presupposition failure, semantic incompleteness, vagueness, reference failure, and the like. If the possession of the intermediate truth-status is ineffable in truth-theoretic terms (one can neither assert nor deny that a sentence is gappy), but not ineffable in non-truth-theoretic terms—terms such as 'not assertible, 'not deniable', and the like—then we might as well dispense with truth-theoretic resources altogether and thereby dispense with the thesis of truth-value gaps.[6] The only interesting form of

---

[6] Holton (2000 pp.10-12) recognises this predicament but seeks to re-express the existence of truth-value gaps via the notion of truth-aptitude: to say that there are truth-value gaps is to say that there are some meaningful declarative sentences which are not truth-apt (p.13). His hope is that we can attribute gappiness indirectly by talking about whether or not a sentence is truth-apt. But then he creates the following problem for himself: since a sentence is truth-apt if and only if it is either true or false then a non-truth-apt sentence will be neither true nor false and so to assert that a sentence is not truth-apt will force us to assert the existence of gaps, which, on the weak reading of the **T** and **F** schemas yields a contradiction. Holton then gives a non-standard non-contraposible conditional in an attempt to block this result. But this is a pseudo-problem: for one thing he attributes this account if truth-aptitude to Wright (1992a), but a key element of Wright's book is that truth-aptitude does not require bivalence. Moreover, truth-aptitude has the following portable characterisation: S is truth-apt just in case S says that something is the case. Holton does not tell us why this portable characterisation, nor the more sophisticated characterisation given by Wright, are to be rejected in favour of his own idiosyncratic reading of truth-aptitude. Furthermore, Armour-Garb (forthcoming) has convincingly argued that Holton's version of compatibilism is in fact incompatible with the expressive role of truth—the role which allows a speaker to

compatibilism concerns the strong reading of truth (from now on I shall take compatibilism to presuppose a strong notion of truth). So let's turn to Horwich's arguments for incompatibilism

Horwich (1990, p. 80) has argued that truth-value gaps issue in contradiction in the presence of certain minimal constraints on our understanding of falsity and negation. He gives the following argument to this effect: If we suppose that <P> is neither true nor false, then <P> is both not true and not false.[7] But given that <P> is false iff <P> is not true, then by contraposition we can infer that if <P> is not false then <P> is not not true; since, by hypothesis, <P> is not false we can therefore derive the contradictory result that <P> is not true and <P> is not not true. The key principle in this argument is what Horwich calls the minimal principle of falsity:

(**MPF**)   <P> is false iff <P> is not true

Given the validity of the biconditional <not P> is true iff <P> is false (i.e., classical negation) and **MPF**, one can derive the biconditional <not P> is true iff <P> is not true (i.e., the recursion clause for negation). But the devotee of truth-value gaps will simply point out that the right-to-left direction of the recursion clause for negation is inconsistent with the existence of truth-value gaps. To reject the right-to-left direction is thus to reject the right-to-left direction of **MPF**. So why must one endorse this direction of **MPF**? Horwich concedes that while one might indeed seek to reject **MPF** and find room for two sorts of falsity— narrow falsity (classical falsity) and broad falsity (non-truth)—he nonetheless thinks that **MPF** has *independent* plausibility. He gives four reasons in all (which I reproduce in full) for accepting the classical account of falsity (Horwich 1998a, p. 77):

> (a) The account reflects our pre-theoretical intuition that if a proposition is *not* true then it is false, and that if something is not the case then the claim that it is the case would be false.
> b) No reasonably plausible alternative characterisation of falsity is able to accommodate these features of the concept.

[7] I here use Horwich's convention of letting '<P>' abbreviate 'the proposition that *P*'. Note that nothing here turns on taking the primary bearers of truth-values to be propositions, rather than declarative sentences— Horwich's argument can just as well be run in the formal mode of speech.

c) The minimalist picture of truth encourages a parallel account of falsity [namely that <P> is false iff not-P] according to which its attributions are similarly equivalent to non-semantic propositions.

d) The spirit of minimalism precludes accounts of truth and falsity which would equip them for theoretical work in semantics.

None of the claims is sufficient to rule out a rejection of the right-to-left direction of MPF within a minimal (or non-minimal) framework. With respect to (d), there is no reason why the spirit of *Horwich's* version of minimalism (roughly, the view that truth is merely some logical property which can play no explanatory role) should move the devotee of truth-value gaps: truth-value gaps are specifically introduced for the theoretical advantages they may bring to semantics. To admit truth-value gaps is just to reject the deflationary (non-explanatory) version of minimalism which Horwich espouses. Admittedly, minimalism is typically conceived to entail that truth should play no explanatory role, but that is just the conception which is up for question. (Our project is methodological: we want to know just which form of minimalism is the best.)

With respect to (c), it depends on what is meant by equivalence. Standardly understood, an equivalence relation is reflexive, symmetric, and transitive. The thesis of truth-value gaps is compatible with the thesis that '<P> is false' and 'not-P' are equivalent in just this sense; likewise this thesis is compatible with the equivalence of '<P> is true' and 'P'.[8] One might seek to give a non-standard definition of equivalence as a relation which is reflexive, symmetric, transitive, *and* contraposible. On this stronger, non-standard, reading of equivalence then there is no scope to reject the right-to-left direction of the recursion clause for negation: <not P> is true iff <P> is not true. In other word, this stronger reading licenses the substitution of '<P> is true' and 'P' within the scope of negation.[9] But to assume a contraposible equivalence relation is just to beg the question against the thesis of truth-value gaps. Hence, Horwich has failed to meet his promise to give an independent argument for **MPF**.

With respect to (a), an appeal to pre-theoretical intuitions is irrelevant. Why should we not expect a theory of truth to accommodate unexpected data? Some 'extraordinary' questions, for instance, can be given no straightforward answer. Here's a very familiar story: If I pose

---

[8] There is the further issue as to whether this equivalence is extensional or modal. Horwich declines to interpret the equivalence schema in modal terms (see 1998a, p. 21, fn.5).

[9] In Dummett's terminology, they would coincide in both content sense and ingredient sense. If the equivalence schema is a necessary truth then it licenses substitution within modal contexts. If this equivalence expresses a coincidence in intensional meaning (whatever that is) then it license substitution in intensional contexts.

the question 'Have you stopped taking drugs?', where it is a fact that the person addressed has never taken drugs, then it is illegitimate for this person to answer *yes* and illegitimate for this person to answer *no*. On a bi-partite reading of the speech acts of assertion and denial, to deny the claim that one has stopped taking drugs, is just to assert that one has not stopped taking drugs. But such a bi-partite model is not exhaustive, for one is also free to *reject* the claim that one has not stopped taking drugs without being committed to assert that one has not stopped taking drugs. Enter truth-value gaps. Given the familiar 'gappy' account of presupposition, it is neither true nor false that I have stopped taking drugs. This possibility makes room for a speech act of rejection (i.e. the speech act of refusing to assert) such that one can reject the claim that *p* without thereby asserting that not-*p*. It is explanatorily advantageous to represent non-truth and falsity as distinct, for in so doing one can thereby ground a tripartite distinction between assertion, rejection, and denial.[10] (It is noteworthy that while Horwich is prepared to consider cases of reference failure, which he hopes to treat via Russell's theory of definite descriptions, (see 1998a p. 78), he omits any discussion of presupposition failure as a reason to recognise truth-value gaps.) Thus, our naive impulse to say that if *S* is not-true then *S* is false provides no grounds to endorse **MPF**.

But what then of Horwich's claim that 'if something is not the case then the claim that it is the case would be false'? That's just seems to be a re-iteration of (part of) the Aristotelian truth-dictum encountered above. If all this claim amounts to is the claim that given that an utterance *u* which says that *p*, what it takes for *u* to be false is just for it to be the case that not-*p*, then we can accommodate this claim by ensuring that the schema **F** be read as a non-contraposible equivalence. For all Horwich has said, such a reading has not been ruled out on independent grounds.

Horwich's strongest challenge is (b). While there's no reason to think that a rejection of **MPF** deprives oneself of an understanding of falsity and negation, it is not a trivial matter to attempt to define negation within some gappy logic. There is a conspicuous burden of proof on the gappist to ensure that the proof theory for a gappy logic does not leave negation (or any correlate falsum constant) undefined. Part of the business of the next section is discharge this burden of proof.

Are there any better arguments which might show that the schemas **T** and **F**, and the schemas **EST** and **ESF**, are to be read as strongly equivalent (i.e. contraposible)? One way to argue for this point is to say that **T** and **F** record an identity of content: either side of these

---

[10] See Parsons (1984) for one of the best discussions of such a tripartite analysis.

119

equivalences are 'same-saying'. In the case of the equivalence schemas **EST** and **ESF**, the thought is that the proposition that is expressed by the words 'it is true that sea-water is salt' is just the same proposition that is expressed by the words 'sea-water is salt'. The earliest remarks to this affect are attributable to Frege, and have served as the foundation for the redundancy theory of truth, whereby, roughly, to prefix 'it is true that' to some sentence does not produce a sentence which differs in meaning from the original.[11] However, Frege admitted truth-value gaps in order to account for terms which are not completely defined for all cases (see his 1903, p. 65). The tension is obvious; and Frege was himself aware of it when he asks: 'How is it then that this word 'true', though it seems devoid of content, cannot be dispensed with?' (1979, p. 252). His answer shows that Frege does not after all accept that truth is redundant in all contexts; he says: 'That we cannot do so is due to the imperfections of language'—where semantic incompleteness is the paradigm example of a linguistic defect for Frege.[12] But the fact that Frege was a compatibilist of sorts is hardly an argument in favour of truth-value gaps. An apparently better argument for an equivalence of content comes from reflecting on the so-called transparency property of truth.

To assert, hope, suppose, doubt, judge, deny, fear that p is to just to assert, hope, suppose, doubt, judge, deny, fear that *it is true that* p, and vice versa. Such is the so-called transparency platitude (see Kalderon 1997). As Wright puts it, 'p' and 'it is true that p' are 'attitudinally equivalent' (2000, p.62). Wright calls this platitude the 'master platitude'. Indeed, it seems right to say that it is this insight which enforces the view that either side of the schema **T** and **F** (and **EST** and **ESF**) are equivalent in content.[13] But is this an equivalence of content which precludes a non-contraposible meaning of these schemas? If one accepts the transparency platitude in full generality, then arguably the answer is *yes*.

Here's how the argument goes: to deny that p is just to deny that it is true that p (and vice versa). But the following is also surely a platitude: to deny that p is to accept that not-p (and vice versa) from which it follows that to deny that it is true that p is just to accept that it is not true that p (and vice versa). But then it follows from these platitudes that to accept that not-p

---

[11] Frege (1918), (1979, p. 251). See also Ramsey (1927), Ayer (1935). With the disquotational theory of truth given by Quine (1970), the predication of 'is true' to some sentence does not entail that we are now talking about the properties of sentences: in asserting 'snow is white' is true I am asserting that snow is white.

[12] Herzberger (1970) has argued that once Frege had drawn the distinction between sense and reference then the horizontal stroke '—' comes to mean something like 'the true proposition' which functions in pretty much the same way as the sentence functor 'It is true that'. Since all the propositional contents in the formal language of the *Grundgesetze* are prefixed with the horizontal, and since Frege seems to have been working with a strong notion of truth, then this has the effect that that any imperfections of language (which give rise to truth-value gaps) can be accommodated: '—A' is either true or false.

is just to accept that it is not true that p (and vice versa). Likewise, for the correlate speech acts of assertion and supposition: to assert/suppose that not-p is just to assert/suppose that it is not true that p (and vice versa; where here we identify speech acts not by the form of words uttered but by the content of what is uttered). But this automatically licenses the inference from $\ulcorner$it is not true that $p\urcorner$ to $\ulcorner$not-$p\urcorner$, which given the uncontroversial inference from $\ulcorner$not-$p\urcorner$ to $\ulcorner$it is false that $p\urcorner$, entails that from $\ulcorner$it is not true that $p\urcorner$ one can infer that $\ulcorner$it is false that $p\urcorner$, which is to say that Horwich's principle **MPF** is validated. Since truth-value gaps are incompatible with **MPF** then the validity of the transparency platitude straightforwardly rules out truth-value gaps. Indeed, since it is uncontroversial that from $\ulcorner$not-$p\urcorner$ one can infer $\ulcorner$it is true that not-$p\urcorner$, then it follows that from $\ulcorner$it is not true that $p\urcorner$ we can infer $\ulcorner$it is true that not-$p\urcorner$, which is just to show that the equivalence schema **EST** is contraposible; ditto for the schemas **ESF**, **T** and **F**.[14] This seems to be the strongest incompatibilist argument available, but is it cogent?

Let's return to our familiar story. Suppose someone asks you: 'is the sentence "you have stopped taking drugs" a true sentence?' To answer 'No', is to deny that this sentence is true, which is to deny that it is true that you have stopped taking drugs which, given the transparency platitude, is to deny that you have stopped taking drugs. But this is just to accept that you have not stopped taking drugs which entails that you are continuing to take drugs. But in answering 'No' to the original sentence one did not intend to imply that one was continuing to take drugs. That's a well-known line of thought, of course, it's just that to take it seriously shows that either the transparency platitude is not valid for all attitudes or the claim that to deny that p is just to accept not-p must be rejected. It doesn't seem an option to reject the right disjunct. Thus, transparency is not valid for the attitude of denial (and cognate attitudes).[15] Note that this argument proceeds independently of whether or not there are truth-value gaps: it has not been claimed that if there are no truth-value gaps then the transparency platitude is true. We have merely removed one conspicuous route to the rejection of truth-value gaps: while the conditional 'if transparency is universally true then are no truth-value gaps' is valid, we have found reasons not to accept the antecedent.

---

[13] If one does not take the transparency platitude to enforce an equivalence of content then at the very least one has to take this platitude to enforce a contraposible equivalence (the argument for this appears below).

[14] Insofar as one accepts that semantic equivalence is to be cashed out in terms of conceptual or inferential roles (as Field 1994a, for instance, does) then this just establishes that the **T** and **F** record a same-saying equivalence.

[15] Presumably, transparency is valid for the attitude of rejection, but to reject that p is not to assert that not-p, and so if the answer 'No' takes the form of a rejection then nothing untoward follows.

But then it looks as if incompatibilism is sanctioned. That's not a conclusion which is mandated however. Arguably, the proper conclusion should be that the transparency platitude is only apparently platitudinous *in full generality*. Indeed, the considerations just given which show that transparency is not universally valid surely chime with our ordinary way of thinking about truth, assertion, and denial. (Just ask the man on the Clapham Omnibus if he has stopped taking drugs.) Truth-value gaps are compatible with a restricted version of the transparency platitude, while the unrestricted version is not only far from being obviously true—it is clearly false for the attitude/act of denial. The minimalist may employ a theory of truth-value gaps with impunity.

4.3 Truth and proof for truth-value gaps

To deny that it is true that p is not to (and does not entail that one must) accept that not-*p*, but to deny that it is true that *p* is to (and does entail that one must) accept that it is not true that *p*. While to accept that *p* is to (and does entail that one must) accept that it is true that *p*. How can we give a logic which parallels, at the level of the truth-theoretic relationships which hold between sentences, these facts about commitments? One way in which one can do so is to have a logic for truth-value gaps in which the right-to-left directions of the schemas **T** and **F**, and the schemas **EST** and **ESF**, are not (classically) contraposible. To help us do this it is useful to borrow a technique developed by Prawitz (1965, pp. 76-8) which allows us to distinguish the 'modal formula' of the language from the unmodalised formula—the relevant modality here being 'It is true that', which we abbreviate with 'T'. (For convenience, we can ignore the predicate modality 'is true'—it's easy to see how the formal results which follow generalise to the formal mode of speech.) Say that: the wff TA, $\sim TA$ are modal; and that the compounds A & B, A v B, A $\supset$ B are modal if their components are. When at least one but not all components of a formula are modal say that the formula is 'mixed' (the term is due to von Wright 1986, 1987). (We also have the modality 'It is false that', abbreviated by 'F' which is definable in terms of T and $\sim$ as follows: FA if and only if $T\sim A$.)

We recognise three truth-values: {True (T), False (F), Neither True nor False (N)}. Despite the fact that we admit a third truth-value, the classical inference patterns hold for the modalised fragment of the language. This is a direct consequence of the semantics of strong truth: V(It is true that A) = T if V(A) = T, = F, otherwise (Smiley 1960: 128; von Wright

1986, 1987). As we shall see, not all classical inference patterns are valid when the language is 'mixed', i.e. when the propositional variables range over both modal and unmodal formula. Furthermore, not all classical inferences are valid when the propositional variables range exclusively over unmodalised formula. (As we shall see, the sequent A → ~A ⊢ ~A, where A is unmodal, turns out to be a case in point.)

Firstly, we can give a restricted rule for the introduction of the classical material conditional, (which we abbreviate with '⊃').

From ⌜Γ, A ⊢ B⌝ infer ⌜Γ ⊢ A ⊃ B⌝   (where A and all members of Γ are either all modal or all unmodal)

When the language is mixed, then we can also give an 'impure' rule (in Dummett's sense) for the introduction of this conditional as follows:

From ⌜Γ, A ⊢ B⌝ infer ⌜Γ ⊢ TA ⊃ B⌝   (where A or some element of Γ may be unmodal) [16]

Neither of these rules allows us to prove the conditional 'If A then it is true that A'. (This conditional is a prototypical mixed formula.) To do that we need to introduce a weaker conditional '→' which sustains detachment but which is non-contraposible. This has an introduction rule as follows:

From ⌜Γ, A ⊢ B⌝ infer ⌜Γ ⊢ A → B⌝   (where A may be unmodal)

This rule holds no matter which formula the variables range over. The corresponding biconditionals ≡ and ↔ are given by the following definitions A ≡ B =$_{df}$ (A ⊃ B) & (B ⊃ A) and A ↔ B =$_{df}$ (A → B) & (B → A). Let the proof-theory for the connectives & and v be just that of classical logic—a point which I will return to in due course.

Given the rule of *T*-introduction which allows us to infer ⌜It is true that A⌝ from ⌜A⌝, and the rule of *T*-elimination which allows us to infer ⌜A⌝ from ⌜It is true that A⌝, then the schema *TA* ↔ *A* is unrestrictedly provable. Given the rule of *F*-introduction which allows us to infer ⌜It is false that A⌝ from ⌜~A⌝, and the rule of *F*-elimination which allows us to infer ⌜~A⌝ from

---

[16] This rule is structurally similar to the to the specification of an 'impure' version of the deduction theorem given by Fine (1975, p. 295): '*B* is a consequence of *A* if and only if *DA* ⊃ *B* is valid'. See also Keefe (2000, pp. 179-180). I should add that this section is based on a paper presented at Logica '99, and written in late 1998, so any overlap between the proof-theoretical results here and those given in Keefe (2000, pp. 175-81) is entirely coincidental.

⌜It is false that A⌝, then the schema $FA \leftrightarrow \sim A$ is also unrestrictedly provable. Given the rules **T1**, **T2**, **F1**, and **F2**, given above then one can likewise prove that S is true $\leftrightarrow p$, and S is false $\leftrightarrow$ not-p. But what of the semantics for the unary connectives? We can give these as follows:

p	$\sim p$	Tp	Fp
T	F	T	F
N	N	F	F
F	T	F	T

This matrix ensures that T and F codify a strong notion of truth. For the binary connectives we have:

$p \; q$	$p \supset q$	$p \equiv q$	$p \, \& \, q$	$p \lor q$	$p \rightarrow q$	$p \leftrightarrow q$
T T	T	T	T	T	T	T
T N	N	N	N	T	N	N
T F	F	F	F	T	F	F
N T	T	N	N	T	T	N
N N	T	T	N	N	T	T
N F	N	N	F	N	T *	T **
F T	T	F	F	T	T	F
F N	T	N	F	N	T	T **
F F	T	T	F	F	T	T

The first four columns of this latter table are simply the familiar three-valued tables given by Lukasiewicz (1930). Note that the $\supset$-I rule ensures that the law of identity $A \supset A$ is provable even for the non-modal fragment of the language. This law is not valid in either Kleene's weak or strong systems, and consequently the deduction theorem fails in those systems since $A \vdash A$ does hold. Since the deduction theorem is an absolutely basic logical principle, then this is a very good reason to reject the credentials of Kleene's systems (at least when the designated value is T and the intermediate value is N).[17] Note also that the connectives & and

---

[17] Kleene intended the intermediate value to be read as something like 'undecided'. Arguably, his tables are only useful for 'irreflexive' purposes, i.e. when one has grounds to reject the identity sequent $A \vdash A$ on the same grounds that one rejects the identity conditional $A \rightarrow A$. When the designated value is 'warranted assertibility' then the liar paradox forces us to reject the identity sequent and as a result the rule of assumptions is not universally valid. See the next chapter. In the context of vagueness, Tye (1990, 1994) employs Kleene's three-valued tables to solve the sorites. Roughly, the thought is that the induction step **A2** is invalid when both antecedent and consequent are neither true nor false. Apart from worries over the deduction theorem, this proposal is unworkable on the grounds that if falsifies such conditionals as 'If a bath of $n°$ is hot then a bath of n+1° is hot' (see Sainsbury and Williamson 1997, p. 478).

v are truth-functional, thus the penumbral connections cannot be expressed using this three-valued logic (again, I shall return to this issue).

Of most interest, is the difference between '⊃' and '→' (marked by *). The conditional → encodes the thought that we are only concerned with whether truth is preserved from antecedent to consequent: if truth is not so preserved (i.e. there is a drop in truth-value from T to N or F) then the conditional is invalid (not T on all assignments) otherwise the conditional is valid (T on all assignments). Given the deduction theorem for →, which we now give as

⊢ A → B if and only if A ⊢ B, this is just to say that an argument is invalid just in case there is an assignment of truth-values under which its premises are true and its conclusion not-true, valid otherwise. That just reflects the fact that *T*-introduction is valid even when applied to a premise which is gappy. But notice that conditional ⊃ is invalid in all cases where there is a 'drop' in truth-value from antecedent to consequent and valid in all other cases. That's just why A ⊃ *T*A is not valid when A is gappy. Thus the deduction theorem, given as ⊢ A ⊃ B if and only if A ⊢ B, is not valid (for mixed formula) since we do have A ⊢ *T*A. Furthermore, while A ⊃ B ⊢ ~B ⊃ ~A, it is should be noted that A → B ⊬ ~B → ~A, and likewise A → B, ~B ⊬ ~A. In other words the conditional '→' does not preserve falsity from antecedent to consequent—this conditional does not support *modus tollens* nor contraposition.

But should we retain two conditionals in a language which may admit of truth-value gaps or should we privilege one over the other? The fact that for → the deduction theorem is not threatened and that one can express the mutual material implication of '*T*A' and 'A' tells in favour of →. Indeed, on very general grounds, why would one want a drop in truth-value from gappy antecedent to false consequent to tell against a three-valued conditional? Truth is the property we hope to preserve in valid inference, not degree of truth, or degree of acceptability, since in supposing a sentence we suppose it to be *true*—we are interested in what follows from the truth of an assumption not what follows from its being neither true nor false. (Should we be interested in what follows from the fact that a sentence is gappy then we have merely to assume *that* the sentence is neither true nor false and see what follows.) Since → is by far the more plausible conditional for a gappy language (which contains mixed formula) then we have reason to entirely dispense with ⊃. Furthermore, it may not be apparent from the syntax alone that one is reasoning in a mixed language. I may believe what Susan said (without really knowing quite what it was that she said) and on that basis believe

125

that something Susan said is true. If I later come to believe that nothing Susan says is true, I am committed to reject my belief, but I am not entitled to assert that this belief is false. Such cases, and countless like them, require us to be cautious, since we are given no guarantee that what Susan said is either true or false. Lastly, it may also be that we require → to express the material (but non-contraposible) equivalence of 'A' and 'It is warrantedly assertible that A'. There may be many other meta-linguistic locutions for which → is valuable (see Williamson 1994, pp.154-6).

What, then, of the difference between ≡ and ↔ (marked by ** in the above matrix)? Very roughly, the rationale behind ↔ is that a biconditional is invalid is there is no parity of truth between both sides, and valid otherwise. In testing for invalidity, we are not interested in cases where there is a disparity in non-truth.[18] It is just these features which ensure that we can express the schemas **EST**, **ESF**, **T** and **F** as a non-contraposible equivalence relation (the above matrices show that A ↔ B ⊬ ~A ↔ ~B). Thus, Dummett's oft-quoted observation that the left-hand side of **T** is false when the right-hand side is gappy, while correct, does not invalidate **T**.

In general, two wff of the language, A and B, are substitutable *salva veritate* (in extensional contexts) only if A ↔ B and ~A ↔ ~B (i.e. where A ≡ B). Since this condition is not met when A is gappy, then that just shows why the substitution step in Williamson's proof is invalid. Indeed, when the connective 'if and only if' in **T** and **F** **is** taken to be non-contraposible this in no way undermines the intuitive Aristotelian rationale for the validity of these schemas. Moreover, we have already seen that the transparency property of truth is not universally valid, whether or not one takes there to be truth-value gaps, and so a non-contraposible reading of these schemas is not precluded on a minimalist conception of truth. It now ought to be clear why the mutual entailment reading of **T** and **F** is a red-herring: there's no reason to reject the idea that **T** and **F** bear a *material* equivalence relation, at least so long as this relation does not sustain classical contraposition. Thus, when ↔ is the only biconditional for the mixed language, a logic for truth-value gaps can have a respectable truth-theory, contrary to what Williamson and Horwich have argued.

However, some of our proof-theoretic problems remain. But these are solvable if we recognise that the conditional → sustains a weaker form of *modus tollens* and contraposition

---

[18] Rather than introduce a new biconditional into the language, Wright (1992, ppo. 63-4) introduces a weak notion of validity in a addition to the strong notion, whereby a biconditional 'A if and only if B' is weakly valid if, although A may in certain circumstances receive a different valuation

such that non-truth is preserved upwards from consequent to antecedent of →. Recall that we already have good reason to think that to deny that it is true that p is not thereby to (be committed to) accept that not-*p*. Of course, a gappy proof-theory cannot validate the *classical* versions of *modus tollens* and contraposition. To demand that they do, would be to beg the question against the thesis of truth-value gaps. It is notable that Williamson (1994, p.152, last paragraph) does not disambiguate informal specifications of contraposition, *reductio*, and *modus tollens* from the classical formal expressions. Thus, he give us no argument that the informal expression of these rules cannot be given an equally plausible non-classical formal specification.

To substantiate these thoughts, we need to reconsider Machina's proof, given in §4.3 above, that truth-value gaps are inconsistent with the rule of negation of introduction. What I take to be the correct response to this argument is in fact anticipated by Machina himself (*cf.* Parsons 1984, p. 142; Priest 1987, p. 16):

> [a gappy approach] cannot allow the validity of indirect proof in general. On that approach, a valid argument leading to a contradiction does lead to a clearly false conclusion, but that merely shows not all the premises in the argument are true. It may be the untrue premises aren't false, either, for they may lack a value entirely. If they do lack a truth value, so do their negations. Hence, one ought not to infer [not-*p* at line (5)*] as we did (Machina 1976, p. 53).

Despite its plausibility, Machina remains entirely unpersuaded that this thought is correct on the grounds that 'whenever a set of premises, *S*, together with some additional proposition, ϕ, has as logical consequence two completely incompatible propositions, ψ and Δ, then *S* has ~ϕ as a logical consequence' (1976, p. 53). If this rationale is right then the thesis of truth-value gaps is indeed inconsistent. But this rational simply begs the question against the thesis of truth-value gaps. Machina is simply not entitled to demand that a logic for truth-value gaps meet this rationale. He *is* entitled to demand that such a logic meet the following rationale: whenever a set of premises, *S*, together with some additional proposition, ϕ, has as logical consequence two completely incompatible propositions, ψ and Δ, then *S* has ~*T*ϕ as a logical consequence. There is no *classical* difference between this rationale and Machina's, thus Machina cannot privilege his rationale over this one. Since the gappist can meet this latter rationale, Machina ought to allow that a gappy approach can allow the validity of indirect proof. So how can we specify the proof-theory in such a way that it sanctions the matrices given above?

Firstly, recall that a sequent is invalid if there is some assignment of values for which the premises are true and the conclusion not true (N or F); and valid otherwise. Below is a selection of the valid and invalid sequents in the language. Given that we are here only dealing with propositional logic, we should expect validity ($\models$) and provability ($\vdash$) to co-vary. Thus, we should expect the sequents below to be provable and not provable respectively:

Valid	Invalid
$p \ \& \sim p \models \perp$	$\models \sim(p \ \& \sim p)$
$\sim(p \lor \sim p) \models \perp$	$\models p \lor \sim p$
$p \rightarrow \sim p, p \models \perp$	$p \rightarrow \sim p \models \sim p$
$\sim p \rightarrow p, \sim p \models \perp$	$\sim p \rightarrow p \models p$
$\perp \models p$	$p \leftrightarrow \sim p \models \perp$
$\sim p \lor q \models p \rightarrow q$	$p \rightarrow q \models \sim p \lor q$
$p \lor q, \sim p \models q$	
$p, \sim(p \ \& \ q) \models \sim q$	
de Morgan's laws	

Table 1.

Take the sequent $p \leftrightarrow \sim p \models \perp$. Given our matrices, when p is gappy the formula $p \leftrightarrow \sim p$ is true, but the falsum constant, $\perp$ is necessarily not-true (by definition). Hence, this sequent is invalid. Equally, we want to ensure that classical inferences patterns hold when the formula to which they are applied are all modal (again that is just a consequence of the semantics for T and F.) In the table below, I give some of the valid and invalid modal and mixed formula, and which ought to be provable and not provable respectively:

Valid	Invalid
$\vDash \sim T(p \mathbin{\&} \sim p)$	$\vDash \sim(\sim Tp \mathbin{\&} \sim T\sim p)$
$\vDash Tp \vee \sim Tp$	$\vDash Tp \vee T\sim p$
$T\sim p \vDash \sim Tp$	$\sim Tp \vDash T\sim p$
$p \vDash \sim T\sim p$	$\sim T\sim p \vDash p$
$\sim p \vDash \sim Tp$	$\sim Tp \vDash \sim p$
$\sim T\sim Tp \vDash Tp$	
$Tp \vDash TTp$	
$\sim TTp \vDash Tp$	

Table 2.

Given Table 2, it looks like the following *reductio*-type rule ought to be generally valid:

From $\ulcorner \Gamma, A \vdash \bot \urcorner$ infer $\ulcorner \Gamma \vdash \sim TA \urcorner$ (Falsum-elimination)

This rule is unrestricted and codifies the idea that the derivation of $\bot$ at most legitimates the inference that some member of the premise set is *not true*, given the possibility that some member of the premise set is neither true nor false. One can also treat this rule as an elimination rule for the absurdity constant $\bot$. A corresponding introduction rule for $\bot$, which also serves as a negation-elimination rule, can be given as follows:

From $\ulcorner \Gamma \vdash A$ and $\Pi \vdash \sim A \urcorner$ infer $\ulcorner \Gamma, \Pi \vdash \bot \urcorner$ (Falsum-introduction)

Again, this rule is unrestricted. Effectively, we are taking the constant $\bot$ to typify any absurdity of the form $A \mathbin{\&} \sim A$. This does not mean that we take $\bot$ to typify a sentence which is necessarily false since not all sentences of this form are necessarily false. $\bot$ typifies a sentence which is necessarily not true. (Note that $\bot$ is not a modal formula.) It should be readily seen that the rule of negation-introduction, as it appears in classical logic, is a

129

composite of three rules: the elimination rule for $\perp$; the standard recursion clause for negation i.e., from $\ulcorner {\sim}TA \urcorner$ infer $\ulcorner T{\sim}A \urcorner$; and (and instance of) the rule of truth-elimination, which permits us to infer $\ulcorner {\sim}A \urcorner$ from $\ulcorner T{\sim}A \urcorner$. To reject the recursion clause for negation is to reject the standard negation-introduction rule.

While the rules of falsum-introduction and falsum-elimination are strong enough to prove the formulas in the left-hand column of Table 2, they are not strong enough to prove the formulas in the first column of Table 1. Nor indeed are they strong enough to prove such sequents as $\vDash T{\sim}(Tp\ \&\ {\sim}Tp)$ and $\vDash T{\sim}{\sim}(Tp\ \lor\ {\sim}Tp)$.[19] To prove these latter, we need to restrict the classical rule of negation-introduction in the following way:

From $\ulcorner \Gamma, A \vdash \perp \urcorner$ infer $\ulcorner \Gamma \vdash {\sim}A \urcorner$ (Negation-introduction)

where (i) this rule is valid if A and all members of Γ are modal. Under such a restriction, the falsum constant effectively typifies all contradictions of the form $TA\ \&\ T{\sim}A$. Thus, $\perp$ is necessarily false, and not merely necessarily not-true. This rule ensures that the classical rule of negation-introduction is valid for the modal fragment of the language. It indeed allows us, for example, to show that $\vdash T{\sim}(Tp\ \&\ {\sim}Tp)$, $\vdash T{\sim}{\sim}(Tp\ \lor\ {\sim}Tp)$.

Given the above rules, we can see that Machina's proof (see §4.3) is addressed directly. Since premise (3)* of the argument fails to be modal, then an application of the rule of negation-introduction is impermissible. An application of the elimination rule for falsum (a rule of untruth introduction, so to speak) is however valid, but this simply enables us to infer just what we should expect—that it is not true that p (*cf.* Edgington, 1993, p. 196, fn.7).

How can we specify the restrictions required to render provable the sequents in column one of Table 1, without having the result that the sequents in column two provable? One natural suggestion is to simply demand that one can derive $\ulcorner B \urcorner$ from $\ulcorner A \urcorner$ only if one can derive $\ulcorner TB \urcorner$ from $\ulcorner A \urcorner$. This ensures that one cannot, for example, infer $\ulcorner {\sim}p \urcorner$ from $\ulcorner p \rightarrow {\sim}p \urcorner$,

---

[19] Here's an example. To show $\vdash Tp\ \lor\ {\sim}Tp$, assume that ${\sim}T(Tp\ \lor\ {\sim}Tp)$. Assume also for reductio that p, given truth-introduction we can infer Tp and $\lor$-introduction then gives us $Tp\ \lor\ {\sim}Tp$. Another step of truth-introduction gives us $T(Tp\ \lor\ {\sim}Tp)$. Contradiction. Using falsum-elimination, reject p to infer ${\sim}Tp$. Another step of $\lor$-introduction and a step of truth-introduction gives us $T(Tp\ \lor\ {\sim}Tp)$. Contradiction. A second application of falsum-elimination gives us ${\sim}T{\sim}T(Tp\ \lor\ {\sim}Tp)$. We now need a primitive rule of double untruth-elimination, so to speak, which allows us to infer $\ulcorner A \urcorner$ from $\ulcorner {\sim}T{\sim}TA \urcorner$. Given such a rule we can thus show that $\vdash Tp\ \lor\ {\sim}Tp$.

since one cannot derive $\ulcorner T\!\sim\!p\urcorner$ from $\ulcorner p \to \sim\!p\urcorner$. Thus, what is derivable in the non-modal fragment of the language is dependent on what inferences hold in the mixed language.

Given, falsum-elimination (i.e. untruth-introduction), it should come as no surprise to see a weaker form of contraposition and *modus tollens*, since these rules are derived rules which depend on the rules of falsum-introduction and elimination. The conditional $\to$ accordingly sustains the following rules:

From $\ulcorner \Gamma \vdash \sim\!B\urcorner$ and $\ulcorner \Pi \vdash A \to B\urcorner$ infer $\ulcorner \Gamma, \Pi \vdash \sim\!TA\urcorner$ (*Modus tollens*)

where all the members of Γ and Π are modal we can infer $\ulcorner \Gamma, \Pi \vdash \sim\!A\urcorner$

From $\ulcorner \Gamma \vdash A \to B\urcorner$ and $\ulcorner \Gamma \vdash \sim\!B \to \sim\!TA\urcorner$ (Contraposition)

where all the members of Γ are modal we can infer $\ulcorner \Gamma \vdash \sim\!B \to \sim\!A\urcorner$.[20] The above formalism is merely a sketch of just how the proof-theory for this three-valued logic should go. But it does seem to offer a system which is both proof-theoretically and truth-theoretically well-motivated.

Two questions remain: Is it legitimate to reject the penumbral connections? Is Wright right to think that a gappy conception of vagueness misconceives the nature of a indeterminacy?

4.4 Penumbral connections and Wright's challenge

As Machina has noted, disputes as to whether or not the penumbral connections are valid can easily deteriorate into 'a battle of raw intuitions'.[21] The three-valued (or many-valued) theorist finds it a merit that $\not\models A \lor \sim\!A$ and $\not\models \sim\!(A \,\&\, \sim\!A)$, while the supervaluationist finds this distinctively implausible. It's of little use in pointing out that supervaluational semantics preserves classical intuitions since it's not clear just which is more classical: truth-functionality or the penumbral connections. Even so, the many-valued theorist feels misunderstood: given his matrices, he is not saying that $\models \sim\!(A \lor \sim\!A)$ or $\models \sim\!\sim\!(A \,\&\, \sim\!A)$. The

---

[20] *cf.* Priest's non-contraposible conditional '$\Rightarrow$' (Priest 1987, pp. 109-10).

law of excluded middle and the law of non-contradiction are taken to be 'quasi-tautologies' in that they are never false; (see Tye 1994; Machina 1976, p.). Put another way, the charge is that supervaluation confuses 'A ∨ ~A' with '*TA* ∨ *~TA*'. The latter is indeed plausible (higher-order vagueness aside) while the former is not (see von Wright 1987). But in itself this is hardly persuasive.

Recall, instead, Dummett's original argument for the penumbral connection of exhaustivity: across the red–orange borderline only two classifications are available for the correct description of the object: red, orange. Hence a borderline object must be either red or orange (Dummett 1975). This is the root intuition behind retaining the penumbral connections. But it's a puzzling intuition when combined with the thought that (extensionally) vague sentences lack truth-values. It's hard to understand why the (standard) argument for the penumbral connections does not equally run when the truth-predicate is in play such that this argument thereby serves as an argument for bivalence. Consider a sentence which asserts that *a* is red when *a* is on the red-orange penumbra, (and so, by the original penumbral argument, *a* is either red or not-red.) There are only two candidate truth-values for the truth of this sentence: true, false. Truth and falsity exhaust the possibilities: thus the sentence 'a is red' is either true or false. But if one finds that argument implausible, on the grounds that it begs the question of the exhaustivity of truth and falsity against the case for truth-value gaps, one should just as much find the original penumbral connections argument implausible on the grounds that it begs the question of the exhaustivity of red and orange (and in turn orange and not-orange) against the thesis that the law of excluded middle is not valid (compare Williamson 1994, p. 162).

Let's consider another argument: How does the intuition that penumbral classifications are exhaustive fare when an object is on the green–blue borderline? According to Dummett, it seems, the colours green and blue exhaust the possibilities, so, a borderline object must be either green or blue. But there is a strong inclination to say such an object is *neither* blue nor green. That's just because we have a ready alternative characterisation to hand, namely that the object is turquoise. But then the thought is that in such a case there isn't after all a borderline region between green and blue since for an object to be penumbral for these two colours there must be no intermediate classification which the (allegedly) borderline object (clearly) satisfies. That ought to be granted. However, in many examples of borderline cases there is a candidate intermediate classification which the object does not clearly satisfy and

---

[21] Not every many-valued theorist rejects the penumbral connections (e.g. Sanford 1976, Edgington 1996).

does not clearly not satisfy. If one varies colour along the brightness axis (rather than on the saturation and hue axes of colour) then what appears to be clearly turquoise in the bright case does not appear to be clearly turquoise in the non-bright case. Thus, there are cases where it is not clear that two candidate colours exhaust all the possibilities and so it is not clear that the penumbral connection of exhaustivity holds. But if there is no penumbral connection of exhaustivity, the route to the law of excluded middle (via the penumbral connection of exclusivity) is blocked. With that route blocked, all the supervaluationist has to fall back on is raw intuition.

Indeed, the thought here generalises to the limiting case when we are dealing with the penumbral connection between a predicate and its complement. The root intuition as to why a borderline object a is either F or not F is that these classifications exhaust all the possibilities. An object on the green–not-green borderline is either green or not. But then why take a borderline case of this sort to exhaust all the possibilities? There is a third possibility, namely that it is not true to say that a is green and it is not true to say that a is not green. The sentence 'a is green' is neither true nor false. The reply to that suggestion is that, even if we were to allow such a possibility, then green and not green cannot properly be said to share a common penumbra for there is an intermediate classification which the object clearly satisfies. The counter-reply is then to say that in many examples of borderline cases, the third possibility neither clearly holds nor clearly fails to hold. In the limiting case, it does hold, and there is no common penumbra between a predicate and its complement—the prototypical examples involving artificial stipulations such as 'oldster'. But in most examples of borderline cases, a predicate and its complement are not exhaustive—there are further predicates which, while not clearly applicable are not clearly not applicable. Again, without exhaustivity, the supervaluationist loses the route to the law of excluded middle.[22]

Such thoughts in any case suggest a shift of emphasis in the standard indeterminist characterisation of what it is to be a borderline case. They also provide the materials to answer Wright's challenge that all gappy or three-valued conceptions of vagueness

---

[22] One might think that a truth-value glut conception of vagueness in which vague sentences can take both truth-values in borderline cases is required to undermine the penumbral connections of *exclusivity*. Indeed, one can develop four valued logics in which statements can be either, neither, or both true or false. The thought would then be that there are four sorts of status which compete for the classification of borderline cases. But, intuitively, too much semantic information is no information at all. If a term is stipulated to both apply and not to apply to certain cases then it can hardly be said to have been given a set of application conditions at all. The reason why the penumbral connection of exclusivity is not validated is because while we can infer from 'it is not true that p and it is not true that not-p' to 'it is not true that: p & not-p' (since the non-truth of both conjuncts entails that a conjunction is not true), since truth and negation do not commute one cannot then infer that it is true that not: p & not-p.

misconstrue the indeterminacy that is characteristic of vagueness. It's worth quoting Wright in full:

> try to conceive being borderline as a status consistent with both the polar verdicts: for an item to be a borderline case on the red–orange border is for it to have a status consistent both with being red and with being orange, (so not red), precisely because it is for that item to have a status under which it *has not been determined* whether it is red or not. If it has not been settled whether or not x is F, that cannot amount to x's having a status inconsistent both with being F and with being not-F; if it were, then matters would have been settled after all—x would be neither. [...] Indeterminacy should be conceived as a matter of things having been left *open*—which requires consistency with each of the relevant polar verdicts in each case (1995, p. 139).

Though Wright takes this to specify what is required of an indeterministic account of borderline cases, the moral in fact generalises. It ought to be a minimal requirement on any theory of vagueness that the status taken by borderline cases is such that it remains uncertain just which polar value is taken by the sentence. Indeterminacy, epistemically construed as in the minimal theory of vagueness, should be conceived as a matter of it remaining unclear just which polar value is taken by an (extensionally) vague sentence. (But only in respect of truth-value since one might well allow that a sentence in borderline cases is clearly neither clearly true nor clearly false). If this challenge is cogent, then truth-value gap conceptions, be they supervaluational or not, are ill-suited to account for the sort of indeterminacy we encounter in genuine borderline cases. Is there a response?

The previous discussion suggests that this challenge is compelling, but that it does not rule out a role for truth-value gaps in our account of what it is to be a borderline case. It is indeed a mistake to say that borderline statements are (clearly) neither true nor false in borderline cases for that is a settled truth-status. Take the analogous case: an object (allegedly) on the green–blue borderline which is in fact turquoise and so not green and not blue; but in certain cases our borderline object is neither clearly turquoise nor clearly green nor clearly blue. On that basis, say that a sentence *S* is extensionally vague just in case it neither clearly true, clearly false, nor clearly neither true nor false. There are thus three polar values: true, false, and neither true nor false. It is left unclear (or indeterminate, if you like) just which truth-status they take. Accordingly, there is indeed a penumbral connection of exhaustivity, it is just that it does not hold between orange and not-orange, it holds between these three polar values.

If these observations are correct, they suggest that there is indeed something amiss with standard truth-value gap conceptions of vagueness, as Wright alleges. While it can be granted

that in the limiting cases, a borderline sentence is clearly neither true nor false, it must also be granted that such cases are highly atypical of borderline cases standardly conceived (cf. Williamson (1997a) on how the vagueness of 'red' differs from the vagueness of incomplete stipulations). In more typical cases, it may be less clear that 'neither true nor false' is the value taken by borderline sentences, just as when the verdict 'turquoise' becomes less obviously correct for an object on the green-blue borderline as the colour of the borderline object is less and less bright. It may well be that in the limiting case it appears to speakers that

the object is clearly not turquoise in which the colours green and blue are penumbrally connected over certain borderline regions. Arguably, in the most typical borderline cases, it is not quite right (and not quite wrong) to say that a is F, not quite right (and not quite wrong) to say that a is not F, and not quite right (and not quite wrong) to say that it is neither true nor false that a is F.[23] While these observations are somewhat sketchy, they do indicate how an account of truth-value gaps might figure in a proper understanding of what it is to be a borderline case. If such an account cannot be stabilised, then there is little hope for truth-value gaps to find a place in a theory of vagueness.

So what conclusions have we reached? We have seen that a logic for truth-value gaps is available which is arguably both truth-theoretically and proof-theoretically acceptable. This logic does not validate the penumbral connections—but that was found to be no bad thing. Furthermore, the minimalist (and non-minimalist alike) is free to utilise such a logic in an (indeterminist) account of what it is to be a borderline case. Whether such a logical system could ground a complete response to the puzzle of vagueness is an issue which lies beyond the scope of this thesis. With the exception of Wright's challenge, I have merely been content to rebut the generic problems which beset gappy logics in general. Can the minimalist employ this logic for truth-value gaps in a solution to the liar paradox?

---

[23] With respect to higher-order vagueness the borderline between the cases where S is not (extensionally) vague and S is (extensionally) vague is itself unclear, where extensional vagueness is as defined in the text.

CHAPTER FIVE[1]
TRUTH-MINIMALISM AND THE LIAR

5.1 The Standard Solution

5.2 Bivalence and illegitimate suppositions

5.3 Is the liar sentence meaningful?

5.4 Supposition and assertion: teleology

5.5 Suppositional inaptitude and the supposition test

5.6 Testing the suppositional credentials of liar sentences

5.7 Supposition and assertion: constitutive rules

5.8 The strengthened liar sentence and the revenge problem

5.9 Semantic closure

Minimalists about truth have had very little to say about the liar paradox. Some have thought that there is a deep reason for this silence—truth-minimalism by its very nature simply lacks any of the right resources to combat the paradox (Simmons 1999). In the previous chapter, it was argued that truth-minimalism is at least compatible with the thesis of truth-value gaps. So can the truth-minimalist employ truth-value gaps in order to solve the liar paradox? According to what Parsons (1984) has dubbed the 'Standard Solution' of the liar paradox, a sentence which says of itself that it is false is a sentence which lacks a truth-value. More sophisticated versions of the Standard Solution take such sentences to be neither *definitely* true nor *definitely* false (McGee 1989, 1991; Soames 1999). The advertised goal of all such proposals is to identify a principled reason to refuse to assert that the liar sentence is (definitely) true/false.

In this chapter, it is argued that while the *form* of the Standard Solution is correct, the reasons why a speaker should refuse to assert that the liar sentence is (definitely) true/false have been systematically misidentified hitherto. An alternative solution (one which retains the shape but not the substance of the Standard Solution) is developed based on the insight that it is improper to even *suppose* the liar sentence to have a truth-status (true or not) on the grounds that supposing a liar sentence to be true/not-true essentially defeats the *telos* of supposition in a readily identifiable way. On that basis, one can block the paradox by restricting the *Rule of Assumptions* in Gentzen-style presentations of the sentential sequent-calculus. The first lesson of the liar paradox turns out to be that not all assumptions are for free.

One key feature of this solution is that provides a positive argument for a minimalist conception of truth, indeed more specifically a *deflationary* conception of truth in which truth should play no role in our philosophical theorising. There are good reasons for this. Contra Simmons (1999), it is not that deflationism is worse off than its competitors with respect to solving the liar paradox on the grounds that deflationism cannot employ truth-theoretic resources for substantial philosophical explanations. Rather, it is argued that a deflationary theory of truth is better off than its competitors. It is the bringing to bear of truth-theoretic resources (such as truth-value gaps) which proves to be problematic and ultimately self-defeating. This is to say that the proposal argued for in this chapter is not merely compatible with deflationism, it provides both a positive and novel reason to accept a deflationary conception of truth. The second lesson of the liar paradox is that deflationism offers the best hope of holding the liar at arm's length.

5.1 Minimalism and the Standard Solution

A sentence which says of itself that it is false is a sentence which lacks a truth-value. Such is the key thesis of what Parsons (1984) has dubbed the 'Standard Solution' of the liar paradox.[2] For all its endurance the Standard Solution has proved hard to stabilise. The familiar stumbling block has been the strengthened liar sentence—the sentence which says of itself

---

[2] The Standard Solution seems goes back at least as far back as Bochvar (1939), and has been defended by numerous authors such as Martin (1967), van Fraassen (1968), Skyrms (1970), and by Parsons (1984).

that it is not true.[3] One natural response to the strengthened liar paradox is to strengthen the Standard Solution in some appropriate fashion. The most sophisticated attempt in this general direction has been given by McGee (1989, 1991). The key idea is to draw a distinction between truth and *definite* truth. A sentence which says of itself that it is not true is a sentence which is neither *definitely* true nor *definitely* false. For McGee, sentences of this sort are 'unsettled' in truth-value—the rules which determine their correct usage give 'bizarre and conflicting answers' (1991, p.8). But any strengthened solution of this general type generates its own form of the strengthened liar sentence—the sentence which says of itself that it is not definitely true.[4] The great merit of McGee's proposal is that steps are taken to address this form of the strengthened liar without recourse to an essentially richer metalanguage. Whether this proposal succeeds (and there are serious, but perhaps not insuperable, doubts on that score) is not the immediate concern in this chapter.[5] The real interest of the Standard Solution (in either its simple or strengthened guise) is whether the shape of the strategy invoked in order to combat the paradoxes provides the basis for a successful solution.

The strategic form of the Standard Solution (simple or strengthened) is more or less based on the following rationale: the liar sentence has some characteristically problematic feature (call it the *L*-property). In virtue of this feature, this sentence ought to receive a particular

---

[3] Let the logical form of this sentence be given by the equality SL = 'SL is not true'. If SL lacks a truth value then SL is not true (and not false). Given the rule of *truth-introduction* (i.e. from $\ulcorner\Gamma\vdash\phi\urcorner$ infer $\ulcorner\Gamma\vdash$ 'ϕ' is true$\urcorner$) then we can derive 'SL is not true' is true. By substituting "SL" for "'SL is not true'" we then derive that SL is true—contrary to the claim that SL lacks a truth-value.

[4] Let the logical form of this sentence be given by the equality DL = 'DL is not definitely true'. If DL is unsettled in truth-value then DL is not definitely true (and not definitely false). Given a rule which, following Heck (1993, p. 203), we may call DEF* (i.e. from $\ulcorner\Gamma\vdash\phi\urcorner$ infer $\ulcorner\Gamma\vdash$ 'ϕ' is definitely true$\urcorner$), then we can derive that 'DL is not definitely true' is definitely true. But by substituting "DL" for "'DL is not definitely true'" we then derive that that DL is definitely true—contrary to the claim that DL is unsettled in truth-value.

[5] Heck (1993) has suggested (in connection with using a 'definitely' operator to model higher-order vagueness) that DEF* (see fn.4) is not a valid rule of inference under indirect, i.e. subordinate, proofs such as conditional proof and *reductio ad absurdum*. Might the same ploy be used against insulating the sentence DL from paradox? This suggestion is not unattractive. However it seems that one can nonetheless reconstruct the paradox in the following way: Suppose (for the sake of argument) that DL is definitely true, then by substitution we can infer 'DL is not definitely true' is definitely true. By the rule of what we may term *def-elimination* (i.e. from $\ulcorner\Gamma\vdash$ 'ϕ' is definitely true$\urcorner$ infer $\ulcorner\Gamma\vdash\phi\urcorner$) we can derive that DL is not definitely true which contradicts our original supposition, and hence we can rigorously prove (by negation-elimination) that DL is not definitely true. Crucially, as McGee (1991, p. 221) notes 'whatever we can prove rigorously is definitely true' which is to say that DEF* ought to be valid under the scope of indirect proofs when the premise set Γ is empty. If so, then we can infer that 'DL is not definitely true' is definitely true, and by substitution we can likewise rigorously prove that DL is definitely true. McGee's response to this formulation of the paradox is to reject the validity of the inference rule *def-elimination* under the scope of indirect proofs (McGee, *ibid.*, p. 222). This has the result that one cannot prove the schema: 'ϕ' is definitely true $\supset \phi$ (a schema which is nonetheless provable for Heck). For doubts about the tenability of this restriction see Priest (1994, p. 388). There are also further problems with McGee's proposal. Mills (1995) has argued that McGee is unable to give a convincing interpretation of what it is to be 'unsettled' in truth-value. Relatedly, there are also very general doubts as to whether 'definitely' can bear a non-epistemic sense (Williamson 1994, pp. 194-95).

evaluative property (call it the *E*-property) which in turn requires us both to refuse to perform the speech act of *S*-ing that this sentence is true, and to refuse to perform the speech act of *S*-ing that this sentence is false. Parsons' version of the Standard Solution, for instance, runs as follows: liar sentences are in some way 'defective' (*L*-property), such that they lack a truth-value (*E*-property), such that

> having discovered that a sentence or proposition does not have a truth-value, we want to reject it, *not* to assert a related sentence (its negation) which we also wish to reject (Parsons 1984, p. 144).

The same strategic template is also employed in sophisticated versions of the Standard Solution. For McGee, liar sentences are governed by conflicting rules of application (*L*-property), such that they are neither definitely true nor definitely false (*E*-property), such that of their truth-status one should say '"I do not know," without intending to intimate that there is any fact of the matter there to be known' (p. 218). Soames (1999) has likewise recently employed the same general strategy. For Soames, the rules governing the use of liar sentences are only 'partially defined' (*L*-property), such that liar sentences are neither *determinately* true nor *determinately* false (*E*-property), such that

> there will be no possible grounds for accepting either the claim that the truth predicate applies to them or the claim that it does not. Because of this, both the claim that such sentences are true and the claim that they are not true must be rejected, thereby blocking the usual paradoxical results (Soames 1999, p. 164).

Where, just as with Parsons, Soames takes the speech act of *rejecting* a sentence to be distinct from the speech act of asserting the negation of this sentence. This latter speech act is usually known as the speech act of *denial*—an act we will encounter again below (see Parsons 1984 for a good discussion of the distinction between rejection and denial).

The problem of the strengthened liar, in all its many guises, then becomes: no matter what *E*-property we identify as justifying both the principled refusal to perform the speech act of asserting that the liar sentence is (definitely) true and the principled refusal to perform the speech act of asserting that it is (definitely) not true, this very property (when fully expressible in the language) permits the reinstatement of some form of the paradox. Indeed Soames (pp. 176-181) concedes that his own approach does not in the end have the resources to combat a strengthened liar sentence of the form 'This sentence is not determinately true'.

(Note that what Soames means by *determinate* truth informally coincides with what McGee means by *definite* truth.)

The *shape* of the Standard Solution *feels* right, even though it has proved difficult to correctly identify the *L*-property and *E*-property which will turn the trick without either reintroducing the paradox in some refined form or without recourse to an essentially richer metalanguage. The nub of such a solution is that possession of the *L*-property is an obvious defect of language. The best response to this defect is a principled silence. In this chapter, it will be argued that both the *L*-property and *E*-property together with the particular speech act of *S*-ing that the liar sentence is true/false (a speech act we must refuse to perform) have all been misidentified hitherto. Rather than nominate the liar sentence as neither (definitely) true nor (definitely) false, in the usual manner, it is put forward that it is illegitimate to *suppose* the liar sentence to be true and illegitimate to *suppose* the liar sentence to be false (not-true). Significantly, the *E*-property here identified is not truth-theoretic. Truth does not, and arguably should not, play any substantive role in our dissolution of the liar. In this respect the proposal advanced in this chapter is deflationist. It is often thought that a deflationist theory of truth is more compromised than most with respect to the liar paradox simply because no (substantive) truth-theoretic resources are available on a deflationary view (Simmons 1999 explicitly expresses this view, though it is implicit in many reactions to deflationism). It is my hope to show that just the opposite is the case. It is rather the bringing to bear of truth-theoretic resources (such as truth-value gaps) which proves to be problematic and ultimately self-defeating. This is to say that the proposal argued for in this chapter is not merely compatible with deflationism, it provides both a positive and novel reason to accept a deflationary conception of truth.[6]

If it is (in a sense to be defined below) illegitimate to suppose that liar sentences have a truth-status then which speech act of *S*-ing that the liar sentence is true/not-true should we

---

[6] Horwich (1998a, p.77) rightly recognises that a deflationist cannot strictly employ truth-value gaps for theoretical work in semantics. Field (1992, p.322, fn.1), strangely, is less sure of this. Since both Field (1994b) and Horwich (1998a, p.79) admit a notion of definite/determinate truth in order to account for such pathologies as vagueness, it might then be thought that they ought to readily endorse some version of the sophisticated Standard Solution of the liar paradox (indeed Soames is a deflationist of sorts). However, in his (1994a, p.250, fn.1, fn.2), Field is content to largely ignore the semantic paradoxes. This is remedied in a forthcoming postscript to this paper, where Field proposes a paraconsistent revision of classical logic along the lines given by Priest (1998). Bradley Armour Garb (forthcoming) has independently argued that dialetheism provides the best deflationary response to the liar paradox. In a similar vein, Priest (1999, p.307) has recently argued that '*Honest* deflationism is not only compatible with dialetheism, it leads in its direction'. Horwich (1998a, pp.40-2) in contrast gestures towards solving the liar paradox by demanding that Tarski's T-schema (or the equivalence schema: the proposition that *p* is true iff *p*) be restricted in some appropriate fashion (though Horwich does not detail how such a restriction is to be effected). Truth, for Horwich, is then to be defined by the maximal consistent set of such instances.

refuse to perform upon discovering this feature? A simple rule governing suppositions runs thus: only suppose what it is legitimate to suppose. A corresponding rule runs: refrain from supposing what it is illegitimate to suppose. (These rules will actually turn out to require qualification—but more of that below.) On the basis of this latter rule one ought to refuse to suppose that the liar sentence is true and refuse to suppose that the liar sentence is false (not-true). This is in contrast to the usual formulations of the Standard Solution where the focus is on the speech act of *rejection*, the speech act of refusing to assert. Soames (1999) asks:

> In what sense do we reject these claims? At a minimum, we must not assert them. However there is more to it than that. We must also hold that it would be a mistake to assert them (p. 171).

There is indeed more to it than that: we must also hold that it would be a mistake to even *suppose* such sentences to be true/not-true. Merely to refuse to assert that the liar sentence is true/not-true is itself an insufficient response to the paradox. A speaker may refuse to assert that the liar sentence is true while nonetheless supposing for the sake of argument that it is true. If this speaker does suppose this for the sake of argument, a paradoxical derivation can be given. The focus on the speech act of *rejection* is a red herring. Refusing to assert the liar sentence is a necessary but not a sufficient response to the paradox. In refusing to suppose *P* (on the grounds that it is improper to suppose that *P*) a speaker is committed to refusing to assert *P*, but not conversely. The speech act of refusing to suppose that the liar sentence has a truth-status (true or not), is however both necessary and sufficient to block the paradox, as we shall see.

5.2 Bivalence and illegitimate suppositions

If liar sentences are not legitimately supposable then how does this feature impact upon bivalence? It is familiar that the principle of bivalence receives a strict and a generalised formulation. The former formulation states that every unambiguous sentence which says that something is the case is either true or false; the latter that such sentences are either true or not true. Strict bivalence is nonetheless compatible with the possibility of what we might call *anodyne* truth-value gaps. Sentences which express questions, commands, or exclamations are neither true nor false, but obviously these sentences do not impugn strict (nor generalised) bivalence—they are anodynely gappy. The same goes for well-formed but meaningless

declarative sentences. (A sentence is 'gappy' in the non-anodyne sense when it says that something is the case but lacks a truth-value.) Bivalence (strict or generalised) is only relevant to sentences (or utterances) that represent the world as thus and so (Williamson 1994, pp.187-88). It is tempting to conjecture that every sentence which is not legitimately supposable must thereby be anodynely gappy. But this thought is too hasty. It depends on just *why* a sentence is not legitimately supposable. Meaningless sentences are indeed not legitimately supposable, and of course these sentences are compatible with, but not subject to, both forms of bivalence. Arguably, however, not all sentences which fail to be legitimately supposable are thereby meaningless. The key thesis of this chapter is that liar sentences are both meaningful and not legitimately supposable. (In the next section, considerations are advanced in favour of the left conjunct of this claim, while in §6-8 arguments are given in favour of the right.) Once we make room for such a possibility then a solution to the liar becomes a genuine prospect.

But if liar sentences are meaningful and yet not legitimately supposable, should we then conclude that they thereby satisfy generalised bivalence but not strict bivalence? Such a thought might be driven by reflection on the following conditionals:

(C1) L is true   →   L is legitimately supposable
(C2) L is false   →   L is legitimately supposable

(where L = 'L is not true', and where '→' is the material conditional). Given C1 and C2, together with the key thesis of this chapter, namely, that L (and 'L is not true') are not legitimately supposable (and the validity of *modus tollens*), it follows that liar sentences are gappy in the non-anodyne sense. If this were so, then the proposal in hand would collapse into the simple Standard Solution and would thus fall foul of the strengthened liar paradox. To secure the proposal, we must find grounds for rejecting C1 and C2.

If we (provisionally at least) take seriously the possibility that, in addition to the semantic values *true* and *false*, meaningful declarative sentences can also take the 'intermediate' semantic value *not legitimately supposable*, then under the most natural interpretation C1 and C2 are to be evaluated as not legitimately supposable. In more detail, if we grant (pending further argument below) that the antecedents of C1 and C2 are indeed not legitimately supposable, then the consequents of these conditionals are, accordingly, false. Under all of the most familiar three-valued matrices for the material conditional, i.e. those given by

142

Lukasiewicz (1930), Bochvar (1939), and Kleene (1952), a conditional with a false consequent but an 'intermediate' antecedent, takes the intermediate value. Since C1 and C2 are not legitimately supposable they are not warrantedly assertible—they should not be accepted, and the proposal in hand does not collapse into the Standard Solution.

One key feature of note here is that the contrapositives of C1 and C2 are likewise not legitimately supposable (they have true antecedents but intermediate consequents). This has the result that it is not legitimate to suppose (and so not legitimate to assert) that the intermediate semantic status excludes truth or excludes falsity. But on that basis it then looks tempting to say that a sentence can fail to be legitimately supposable but nonetheless remain either true or false (where falsity for our purposes is equivalent to non-truth). However this is not so. Generalised bivalence is best formulated as a conjunction of two principles:

Principle of valence: Every meaningful declarative statement has a truth-status
Principle of two truth-status: There are two sorts of truth-status: *true, not-true*[7]

The thesis that it is not legitimate to suppose that that the liar sentence has a truth-status, entails that it is not legitimate to suppose (and hence to assert) the *principle of valence*. Since there is no (overt) worry with the *principle of two truth-status*, then on the plausible assumption that a conjunction with one true conjunct and one intermediate conjunct must take the intermediate value, then generalised bivalence is not legitimately supposable and so not legitimately assertible, where crucially, this does not entail that bivalence is deniable—that its negation is assertible.[8] And so, in addition to the triad of positions *anodynely gappy, non-anodynely gappy but not strictly bivalent*, and *non-anodynely gappy but strictly bivalent*, there is a further status which meaningful but non-legitimately supposable sentences may take, namely a status for which all forms of bivalence are themselves not legitimately supposable. We have glimpsed how such a proposal impacts upon classical semantics. Now we must endeavour to secure the thesis that liar sentences are indeed meaningful.

---

[7] Cf. Sanford (1976, p. 196).
[8] It should also be noted that the principle of generalised bi-exclusion (which says that no meaningful declarative sentence is both true and not true) is likewise not legitimately supposable.

To answer this question in detail would require more space than is available here, so what follows is just an outline of how the arguments might run. The immediate evidence strongly suggests that there is no particular reason to doubt that liar sentences are devoid of content. The sentence 'This sentence is not true' is certainly grammatical. Nor would it seem to represent a category mistake, for the right kind of predicate is predicated of the right category of thing. Furthermore, each word would also seem to bear its usual meaning, and there ought to be no particular worry concerning self-reference—just as the (false) sentence 'This sentence contains ten words' says that something is the case, so does the liar sentence. With these observations in mind, it is surprising to find how many authors have thought that liar sentences (or utterances of liar sentences) fail to represent the world as thus and so.[9] At first sight, such a 'no-proposition' view of liar sentences seems susceptible to a version of the strengthened liar paradox. If a liar sentence fails to say that something is the case (i.e., fails, for all intents and purposes, to express a proposition) then it lacks a truth-value by default since it cannot be a *bona fide* truth-bearer. Accordingly, it seems one can run the strengthened liar paradox given in footnote 3 against such a proposal. But this is too quick, for the rule of *truth-introduction* employed there was in fact stated too simply. This rule (and the corresponding rule of *truth-elimination*) should rather be stated, respectively, as follows:

If ϕ says that something is the case, then from $\ulcorner \Gamma \vdash \phi \urcorner$ one can infer $\ulcorner \Gamma \vdash$ 'ϕ' is true$\urcorner$

If ϕ says that something is the case, then from $\ulcorner \Gamma \vdash$ 'ϕ' is true$\urcorner$ one can infer $\ulcorner \Gamma \vdash \phi \urcorner$

These rules ensure that semantic ascent and descent are permitted if it is first given that 'ϕ' says that something is the case (cf. Williamson 1994, pp.187-8). Since the sentence which says of itself that it is not true does not say that something is the case, then the consequent of these rules is not validated and no strengthened liar paradox is derivable. Is this no-proposition response at all cogent?

There are two conspicuous problems with the no-proposition response to the liar paradox. Firstly, it would appear that in any case one can reconstruct the paradox in terms of propositions rather than sentences. Let 'Π' stand for the proposition that Π is not true. Assume that Π is true. Then given what 'Π' stands for, this is just to say that the proposition

---

[9] See e.g. Bar-Hillel (1957) for an early statement of this view.

that Π is not true is itself true. Given the 'equivalence thesis' (i.e., the proposition that Π is true if and only if Π) then we can infer that Π is not true. Contradiction. Conclude (by negation-introduction) that: Π is not true. But given the equivalence thesis we can now infer that the proposition that Π is not true is itself true, and given that 'Π' stands for the proposition that Π is not true this is just to say that Π is true. Paradox.

The second problem turns on the possibility of contingent liar sentences.[10] If I inscribe on my whiteboard the sentence 'Some sentence on this whiteboard is not true', then whether this sentence counts as liar-like depends on the contingent fact as to whether or not there is more than one sentence inscribed on the whiteboard.[11] Suppose I rub out all other sentences bar this one sentence. While we should expect such a change to affect the *E*-property we take this sentence to have, we should not expect any such change to affect whether or not this contingent liar sentence says that something is the case. According to certain versions of the simple Standard Solution, for instance, rubbing out all other sentences on the board will affect whether or not the sentence in hand has a truth-value, but will not affect whether this sentence has truth-conditions. In more neutral terms, changes in the world can affect whether or not a statement is warrantedly assertible, but these changes need not have any direct impact on whether the statement has warranted assertibility conditions. Of course much more could be said about this matter, but there is at least a strong *prima facie* case to think that liar sentences are meaningful.[12]

Thus far nothing has been said as to what sort of *L*-property liar sentences possess which justify the evaluation that it is illegitimate to suppose that such sentences have a truth-status

---

[10] A proposal of this sort also suffers from what has come to be known as the 'revenge problem'. Roughly speaking, any solution to the liar paradox suffers from the revenge problem when it has pathological (but not necessarily inconsistent) consequences. (Generally speaking most authors run the revenge problem together with the problem of strengthened liar paradox, when in fact the former need not entail the latter.) If SL is meaningless such that SL is not true then, given substitution, we can also assert that 'SL is not true' is not true. Since we cannot employ the rule of *truth-introduction*, this falls short of being a proper contradiction. It nonetheless remains pathological since in asserting that SL is not true, a speaker does not thereby assert that 'SL is not true' is true. Hence the Fregean platitude that to assert *that P* is to assert *that it is true that P* would appear to be a casualty of a proposal of this general sort.

[11] The example is adapted from Mackie (1973, p. 294).

[12] The deflationist will in fact argue that the driving thesis behind the *no-proposition* view is the truth-conditional conception of meaning and understanding. The impossibility of truth-evaluating the liar sentence is taken to be evidence that this sentence fails to have truth-conditions and so fails to say that something is the case. But it is well-known that the truth-conditional conception of meaning has unpalatable consequences. Furthermore, alternative theories of meaning and understanding are available in which grasp of meaning does not entail grasp of truth-conditions, and in which truth plays no substantial explanatory role. One can give a theory of content via reference to warranted assertibility conditions, or one might seek to give a use-theoretic, or conceptual role, model of meaning and understanding. This is not the place to defend such deflationary theories of content here—but see Field (1994b) and Horwich (1998b). Arguably, such accounts provide added confirmation that liar sentences do indeed bear content.

(true or not). The claim developed below is that non-contingent liar sentences possess a distinctive logical form—a form which inevitably undermines the *telos* or goal of the speech act of supposition. It proves possible to identify a syntactic (rather than a truth-theoretic) *L*-property of non-contingent liar sentences which dictates that that a speaker must not suppose such sentences to have a truth-status. To secure this claim we must first survey some salient features of the speech act of supposition. To this end it is useful to begin by comparing the speech act of supposition with that of assertion.

5.4 Supposition and assertion: teleology

In what follows, it is merely necessary to uncover those aspects of supposition which are directly relevant to a dissolution of the paradox.[13] First some preliminaries. The term 'supposition' is ambiguous. On the one hand, we can speak of supposition as a species of speech act, and on the other, we can speak of the sentence or proposition which is the object of that speech act. In what follows, it is used to refer to the former. It is also germane to speak of suppositions in a broader sense—as acts of linguistic inscription and as mental acts which an individual can perform without necessarily uttering sounds. One can think of utterances which say that something is the case (i.e., assertions, suppositions, conjectures, etc.) as being the primary bearers of truth-values, or one can think of these acts as bearing truth-values only insofar as they express propositions or have as their objects meaningful declarative sentences. For the sake of convenience, we can take declarative sentences to be the primary truth-bearers, simply because the debate concerning the liar paradox has conventionally dealt with the problems attending liar *sentences*.

Supposition is a goal-directed activity. In supposing, quite simply, we are interested in establishing what follows from what. Supposition in this sense, as we should expect, is governed by teleological norms. Teleological accounts of assertion are familiar from the writings of Dummett (1959, 1973, p. 320; see also Priest 1987, pp. 77-9, 2000, pp. 309-10). The point or goal of assertion, on these accounts, is to utter true sentences—to hit the truth. The teleological norm governing assertion thus runs: in asserting, aim to say what is true. Making assertions for Dummett and Priest is usefully compared with a game: to utter truths is to win, while to utter falsehoods is to lose. Call this the *truth-account* of assertion. (Assertion

146

is here more or less conceived in the Fregean sense as the 'outer' manifestation of the mental act of judgement, an act whose attitudinal correlate is belief. A more refined view might maintain that while the telos of judgement/belief is truth, the telos of assertion is truth *plus* the communication of truth.) A stronger teleological account says that it is constitutive of assertion that the *telos* of assertion is knowledge (Williamson, 2000, p.1, expresses a version of this stronger view by saying that 'the point of belief is knowledge'.) The teleological norm on this account runs: in asserting, aim to say what you know to be true. Call this the *knowledge-account* of assertion. This is not the place to defend this account in detail, but the knowledge account is surely more compelling. Though it's harder to win at the game of assertion on the knowledge account, we do not want to win at this game by accident: our assertions are required to be reliability right—a condition that the truth account cannot enforce.

What then of the telos of supposition? Suppositional reasoning is intimately connected with the categorical assertion of conditional claims. This fact is reflected in the validity of the deduction theorem: $A \vdash B$ if and only if $\vdash A \to B$ (where '$\to$' is the material conditional, and '$A \vdash B$' abbreviates 'B is provable (in some unspecified proof-theory) given A', and where '$\vdash A \to B$' abbreviates '"$A \to B$" is a theorem, i.e. provable on no assumptions'). It makes no sense to speak of *mere* supposition. In supposing some sentence A, one is interested in giving valid proofs of what follows from A. Generally, we suppose some sentence A in order show *whether or not* some sentence B is provable from A. [14] In particular, we aim to be in a position to assert '$\vdash A \to B$' truly or be in a position to assert '$\nvdash A \to B$' truly. So, the teleological norm governing the supposition of A (in order to see whether B follows) runs: aims to be in a position to truly assert that B is provable from A or to be in a position to truly assert that B is *not* provable from A. Call this the *truth-account* of supposition. In contrast, the stronger knowledge account of suppositional reasoning says that in supposing A, one must be in a position to know that B is provable from A or be in a position to know that B is *not* provable from A. To win at supposition, it is not enough for one to truly assert whether or not B follows from A. One loses at supposition, for instance, if one's assertion that $A \vdash B$ is indeed

---

[13] The best work on supposition has been done by Cargile (typescript); Dummett (1973, pp. 309-10); Green (2000); (Kearns (1997), and Kearns (typescript).

[14] There need be no requirement that a speaker must have some particular sentence B in mind. Occasionally we wish to suppose a sentence when not having a very fixed idea of what its putative logical consequences might be. (I am indebted to Patrice Philie for stressing this point.)

true, but where one's belief that $A \vdash B$ could easily have been wrong—one does not want to win at the game of supposition by accident. Again, this is not the place to defend such a knowledge account in detail, but for this reason alone, the knowledge account is surely more cogent. It is crucial to note that the goal of supposition is stated as an *exhaustive* disjunctive condition: aim to either know that $A \vdash B$ or know that $A \nvdash B$, where one fails to satisfy this goal if one is in neither epistemic position.

5.5 Suppositional inaptitude and the supposition test

One may fail to satisfy the point of suppositional reasoning for a variety of reasons. One may fail to be in a position to know that B follows from A or to know that B does not follow from A, simply through limitations on one's powers of logical deduction. Sometimes the very integrity of the supposed sentence is the root reason for failing to win at the game of supposition. This occurs, for instance, in the case of supposing sentences which do not bear a proper content. In supposing the sentence 'Jim is slithy mimsy brillig and generous' (call this A) in order prove whether or not the sentence 'Jim is generous' (call this B) logically follows, one cannot know that $A \vdash B$ or know that $A \nvdash B$ since, even though the inference is *formally* valid, and 'B' is a truth-bearer, the sentence 'A' is plainly gibberish and so not a proper truth-bearer. Here we should rather say that it is not our reasoning that is at fault *per se*, but the very supposition of the sentence which features as antecedent. In this case, there is *in principle* no warrant to accept/deny *all* conditionals in which a meaningless statement features as antecedent—we are not in a position to know that $\vdash A \rightarrow B$, nor in a position to know that $\nvdash A \rightarrow B$. We are thus entitled to say that the speech act of supposing A is *essentially* improper. It thus pays at this point to introduce some terminology to refer to those sentences which may, for whatever reason, essentially defeat the goal of supposition. Say that

> a sentence 'A' fails to be *supposition-apt* if there is *in principle* no warrant for a speaker to accept or deny *all* conditionals in which 'A' is the antecedent.

This effectively characterises what we may call *generic* supposition-inaptness (a more specific characterisation will be given in a moment). A sentence is supposition-apt just in case it is not supposition-inapt (just in case, that is, there is *in principle* some knowledge

148

conferring warrant to accept that $\vdash A \rightarrow B$ or some knowledge conferring warrant that $\nvdash A \rightarrow B$).

Sentences may be supposition-inapt for a variety of reasons. Ungrammatical sentences, sentences which embody category mistakes, nonsensical sentences, and so forth, are all supposition-inapt. These are all sentences which fail to say that something is the case. However, lack of proper content is a sufficient but not a necessary condition of suppositional inaptitude. We should also allow that meaningful sentences may essentially defeat the telos of supposition. One way in which this might occur is when one has *both* a warrant (or reason) to accept that $\vdash A \rightarrow B$ *and* a warrant (or reason) to accept that $\vdash \sim(A \rightarrow B)$ and so a warrant (or reason) to accept that $\nvdash A \rightarrow B$ (given that warrants transmit over the entailment from '$\vdash \sim(A \rightarrow B)$' to '$\nvdash A \rightarrow B$'). Hence, one cannot be in a position to know that $\vdash A \rightarrow B$ since the evidence one has for $\nvdash A \rightarrow B$ (such as a putatively valid proof) defeats the possibility of this knowledge, but nor can one be in a position to know that $\nvdash A \rightarrow B$ since the evidence that one has for $\vdash A \rightarrow B$ (such as a putatively valid proof) defeats the possibility of this knowledge also. One fails to win at the game of supposition in such a case.

Such observations suggest a more specific formulation of suppositional inaptitude which will enable us to isolate the particular *L*-property possessed by liar-sentences in virtue of which they are supposition-inapt. We can do this by submitting the suppositional credentials of declarative sentences to the following test:

The Supposition Test. A sentence '*A*' fails to be supposition-apt if, for *all* sentences '*B*', one can establish that (i) $\vdash_{NK+} A \rightarrow B$, and (ii) $\vdash_{NK+} \sim(A \rightarrow B)$.

First some minor comments on this test: (a) Note that '$A \vdash_{NK+} B$' abbreviates '*B* is provable given *A*, in classical logic (plus the rules of truth-introduction and truth-elimination)'. (b) If a sentence is supposition-inapt in this more specific sense then it will be supposition-inapt in the generic sense defined above, but not vice versa. (c) Strictly speaking, one ought to add a third condition to the effect that the rules of proof in NK+ are beyond reproach (as indeed they are—apart that is, from the rule of assumptions as we shall see). This ensures that when (i) and (ii) are satisfied, we are indeed blaming the suppositional credentials of '*A*' rather than

the system of proof itself. (d) This test is only relevant to non-contingent liar-sentences, for in deriving a paradox from a contingent liar sentence the premise set is never empty—it must contain some relevant contingent assumption. (For the sentence 'some sentence on this page is not true' to be paradoxical, it must depend on the contingent assumption that there is only one sentence on this page.)[15]

But what exactly does this test amount to? Clause (i) says that B is derivable from A, while clause (ii) entails that B is *not* derivable from A. If both clauses are satisfied then clearly something has gone wrong—but what? The obvious response is to blame our proof-theory— that some rule of proof in NK+ does not preserve the designated value and requires restriction in some appropriate fashion. (In fact the culprit is indeed the rule of assumptions, as we shall see below.) More informally, satisfaction of (i) and (ii) shows that we are essentially prevented from finding out what any of the logical consequences of the sentence 'A' are. When the *telos* of supposition is defeated in this absolute we can say that 'A' is supposition-inapt: there's no point in supposing a sentence if one can never be in a position to demonstrate what follows from this sentence.[16] Given the above discussion, we are now in a position to show that contradictions pass the supposition test while liar sentences characteristically do not.

6.6 Testing the suppositional credentials of liar sentences

The most that can be inferred from the supposition of 'P & $\sim P$' is that $\vdash_{NK+}(P\&\sim P) \to P$ and $\vdash_{NK+}(P\&\sim P) \to \sim P$. Clause (ii) is not satisfied, and so contradictions (which themselves contain no liar sentences) pass the supposition test. If we could additionally establish that $\vdash_{NK+}(P\&\sim P)$, then the matter would be different, for then (given *modus ponens*) we could then infer $\vdash_{NK+}\sim P$, and given $\vdash_{NK+}(P\&\sim P)$, this is to show that $\vdash_{NK+}\sim((P\&\sim P)\to P)$, and so clause (ii) would be satisfied also. This is exactly what happens with the liar paradox.

---

[15] A complete treatment of how a proposal of this sort extends to contingent liar sentences is beyond the scope of the present work.

[16] One might be tempted to think that liar sentences are not legitimately supposable on the grounds that in supposing a liar sentence to be true one can show in NK+ that every sentence is derivable (since one can show that $\vdash \perp$ and the *ex falso quodlibet* is valid). But it's not the classical spread principle *per se* which enables us to say that the supposition of liar-sentences is illegitimate. The provability of everything is only relevant once we recognise the teleological norms governing supposition. The provability of everything is a necessary but not a sufficient condition for 'L is true' to count as suppositionally inapt.

However, to show that liar sentences are supposition-inapt in this way we need to firstly address a preliminary puzzle.

Clauses (i) and (ii) are very demanding in the sense that it must be shown that B both is and is not derivable from A, for *all* substitution of the sentential variable B. But surely we should expect a sequent L is true ⊢ ~(Jam is red and Jam is not red) to be a valid sequent, and, given the deduction theorem, we should likewise expect it to be the case that ⊢$_{NK+}$ L is true → ~(Jam is red and Jam is not red).[17] Hence, 'L is true' ought to be able to feature as the antecedent of certain unproblematic conditionals. If it can do so, then 'L is true' passes the supposition test (contrary to the advertised aims of the proposal). However in NK+, *thinning* (i.e. the structural rule of dilution/weakening: *from* $\ulcorner\Gamma\vdash_{NK+}B\urcorner$ *infer* $\ulcorner\Gamma, A\vdash_{NK+}B\urcorner$) is valid, and so liar-susceptibility can be shown to be 'infectious'. For instance, one can show that the supposition that 'L is true v ~(Jam is red & Jam is not red)' gives rise to paradox if one admits thinning.[18] This infectiousness indicates why it is pertinent to ensure that clauses (i) and (ii) hold for every substitution for B.

To show that a liar sentence fails the supposition test we need to move in two stages. Firstly, we need to show that clause (i) and (ii) are satisfied for all the 'relevant' putative consequences of liar sentences; secondly, we need to show that these are satisfied for all the 'irrelevant' putative consequences of such sentences. The distinction between relevant and irrelevant is roughly that intended by relevance logicians as we shall see. A sentence B is a 'putative' logical consequence of B when there is proof-theoretically valid demonstration that B follows from A in NK+. As indicated above this does not mean that B is a *bona fide* logical consequence of A as there may be a proof-theoretically valid demonstration that B does not

---

[17] One reason for this expectation is that these sequents contain what we might call 'mixed' formulas. The sentence 'Jam is red' belongs to the non-semantic fragment of the object language, while 'L is not true belongs to the semantic fragment' (we are here assuming semantic closure—see the last section). On that basis, we are given a syntactical guarantee that 'Jam is red' is not liar-susceptible, and so we should expect it to be subject to the theorems and inferences of classical logic.

[18] Proof: (let 'P' abbreviate 'Jam is red'). Assume L is true v ~(P & ~P), and assume L is true. By &I and &E, one then infers that L is true (which now depends on both assumptions). From there one can derive ⊥ and so, by ~I on the second assumption infer that L is not true, from which one can then derive ⊥. One then uses ~I to reject the first assumption to yield ~(L is true v ~(P & ~P)), and by de Morgan, we can derive L is not true & ~~(P & ~P). From the left conjunct of this we derive ⊥ (which rests on no assumptions). Paradox. If ~(P & ~P) is immune from liar-susceptibility then one would expect L is true v ~(P & ~P) to be likewise immune, but it's not.

151

follow from A.[19] Let's turn to look at a relevant putative logical consequence of the liar sentence.

Let the logical form of the liar sentence be given by the usual equality L = 'L is not true'. Let 'A' stand for the sentence 'L is not true', and let 'B', the putative candidate consequence of A, abbreviate this very same sentence. Then suppose

1 (1)   L is not true

then by conditional-introduction we can straightforwardly show that $\vdash_{NK+} A \to B$:

(2)   L is not true $\to$ L is not true

Clause (i) is satisfied. But we also need to establish $\vdash_{NK+} \sim(A \to B)$. For this it suffices to establish both $\vdash_{NK+} A$ *and* $\vdash_{NK+} \sim B$. Given the rule of truth-introduction, from line (1) we can infer:

1 (3)   'L is not true' is true

and given that L = 'L is not true', then by substitution in (3) we infer

1 (4)   L is true

Contradiction. So, by negation-introduction we can infer:

(5) $\sim$(L is not true)

---

[19] A sentence B is a *bona fide* semantic consequence of A only if (a) $\vdash_{NK+} A \to B$ (b) $\nvdash_{NK+} \sim(A \to B)$. It might be thought that clause (b) is superfluous since it looks like it follows from (a). But that entailment is predicated on the assumption that from the metalinguistic statement '$\vdash_{NK+} A \to B$' one can assert that $A \to B$, and from the metalinguistic statement '$\nvdash_{NK+} \sim(A \to B)$' one can infer that $\sim(A \to B)$ and derive a paradox from which one infers that $\nvdash_{NK+} \sim(A \to B)$. But these inferences are only valid if '$\vdash$' is already factive, so to speak, and to have that property it must already be answerable to semantic consequence which is to beg the very question at issue.

which establishes $\vdash_{NK+}\, \sim B$. By the rule which allows us to infer $\ulcorner$'ϕ' is not true$\urcorner$ from $\ulcorner\sim\phi\urcorner$ we now derive

(6)   'L is not true' is not true

and given that L = 'L is not true', then by substitution in (6) we infer

(7)   L is not true

which establishes that $\vdash_{NK+}A$. Given that '$\rightarrow$' is the material conditional, to establish both $\vdash_{NK+}A$ *and* $\vdash_{NK+}\sim B$ is to establish that $\vdash_{NK+}\sim(A \rightarrow B)$.

The result generalises. Let 'A' represent the sentence "L is not true", and let 'B' represent the sentence "'L is not true' is true". By conditional introduction on lines (1) and (3) we can establish that $\vdash_{NK+}A \rightarrow B$. Given that at line (7) we have $\vdash_{NK+}A$, and at line (6) we have $\vdash_{NK+}\sim B$, then we have again shown that $\vdash_{NK+}\sim(A \rightarrow B)$. Take another example, namely the sequent $\vdash_{NK+}$ L is true $\rightarrow$ (L is true v P). We have already proved that $\vdash_{NK+}$ L is true, and the sequent itself is easily provable. We now need to prove that $\sim$(L is true v P). But since from lines 5 and 7 it in any case follows that from $\vdash\perp$, then by *ex falso quodlibet* we can derive $\vdash_{NK+}\sim$(L is true v P).

Clearly, for any liar-like sentence A, and for any putative relevant consequence of this sentence B, we can always demonstrate that clauses (i) and (ii) are both satisfied—at least if we allow ourselves the full resources of NK+, including the classical spread law. Consequently, liar sentences fail to pass the supposition test, at least for all 'relevant' substitutions for B. For liar sentences to fail to pass the supposition test in full generality, then for *all* substitution for B (be they relevant putative consequences or an irrelevant putative consequences) one must likewise be able to establish that (i) $\vdash_{NK+} A \rightarrow B$, and (ii) $\vdash_{NK+} \sim(A \rightarrow B)$. For instance, in order to ensure that 'L is true is supposition-inapt', we at the very least need to establish that $\vdash_{NK+}$ L is true $\rightarrow B$, $\vdash_{NK+}$ L is true $\rightarrow \sim(B)$, and $\vdash_{NK+}$ L is true, where 'B' ranges over the classical theorems. More than that, we need to let 'B' range over such

irrelevant consequences as 'the moon is made of cheddar cheese', and the like. One can secure this result in a variety of ways, but I shall use the 'paradoxes of strict implication'.

Let A be the sentence 'L is not true' as before. Line (5) effectively establishes that $\vdash_{NK+}\sim A$. Given the *Rule of Necessitation*, we can then infer that $\vdash_{NK+}\Box\sim A$. Given the paradoxes of strict implication, we can then infer both that $\vdash_{NK+}A \prec B$ and $\vdash_{NK+}A \prec \sim B$ (for all B). Since strict implication $\prec$ entails material implication $\rightarrow$ then this is just to establish that $\vdash_{NK+}A \rightarrow B$ and $\vdash_{NK+} A \rightarrow \sim B$. Since we have already shown that $\vdash_{NK+}A$, then by *modus ponens* we can now infer that $\vdash_{NK+}\sim B$, which gives us $\vdash_{NK+}\sim(A \rightarrow B)$, and so both clauses (i) and (ii) are satisfied for any sentence B, be it relevant or not.[20] So, it is illegitimate to suppose a sentence for which we are never in a position to accept or deny what putatively follows (relevantly or irrelevantly) from that sentence. Liar sentences are essentially unfit to enable the telos of supposition to be satisfied. But on that basis of having that L-property, how can one appropriately restrict the proof theory of NK+? To answer that question we must return to our comparative analysis of supposition and assertion.

5.7 Supposition and assertion: constitutive rules

Teleological norms for speech acts do not necessarily coincide with the norms given by what Williamson (1996) has called the *constitutive* rules which govern such acts. These rules codify the norms which *essentially* and *uniquely* govern each speech act.[21] On the Williamsonian model, each and every speech act can be identified by reference to the rules which are constitutive for it, and in turn we can evaluate the performance of a particular speech act on the basis of its constitutive rule. As Williamson (p. 491) puts it:

---

[20] Alternatively one could use *ex falso quodlibet*, i.e., from $\ulcorner\vdash_{NK+}\bot\urcorner$ infer $\ulcorner A\urcorner$. Given that we can use the liar paradox to establish that $\vdash_{NK+} \bot$, then given EFQ we can establish both that $\vdash_{NK+}B$ and $\vdash_{NK+}\sim B$. And by two applications of conditional-introduction on any line in which a liar sentence P is supposed, we can establish that $\vdash_{NK+}A \rightarrow B$ and $\vdash_{NK+}A \rightarrow \sim B$, and since we already have $\vdash_{NK+}B$ then the same result follows.

[21] As Williamson says: 'The normativity of a constitutive rule is not moral or teleological. [T]he criticism that one has broken a constitutive rule of an institution [is not] the criticism that one has used it in a way incompatible with its aim' (1996c, pp. 491-2). For example, I may fail to satisfy the teleological norms which may govern the game of chess (e.g. the norms: aim to win, aim to exercise your mind, aim to pass the time, etc.) without thereby breaking any of the rules which are constitutive of chess.

Constitutive rules do not lay down necessary conditions for performing the constituted act. When one breaks a rule of a game, one does not thereby cease to be playing that game. [...] Likewise, presumably, for a speech act: when one breaks a rule of assertion, one does not thereby fail to make an assertion. One is subject to criticism precisely because one has performed an act for which the rule is constitutive.

In the particular case of assertion, Williamson pursues the 'simple' thesis that there is just one constitutive rule governing this speech act, a rule of the following form:

The C(P) rule.　　One must: assert that *P* only if C(*P*).

Where C(*P*) represents some condition with respect to *P*. This rule serves to forbid the possibility that a speaker asserts *P* when not C(*P*). Williamson considers a variety of candidates for the condition C(*P*)—truth, warrant, knowledge—and settles upon the strong condition that only knowledge warrants assertion. Thus the single rule of assertion is:

The A-rule.　　One must: assert that *P* only if one knows that *P*.

It is clear that a speaker may break the *A-rule* and yet satisfy the telos of assertion (according to the truth-account of assertion) for a speaker may hit the truth in asserting *P* without knowing that *P* is true. Equally, a speaker may respect the *A-rule*, and yet fail to satisfy the telos of assertion on the knowledge account by failing to *communicate* a known truth. (That said, it does seem that there is a more intimate connection between the constitutive rules and teleological norms than Williamson concedes in the quote given in fn. 21).

The evidence Williamson marshals for the correctness of the *A-rule* is compelling. It would be a Moorean paradox to assert '*P* but I do not know that *P*'. The pathology of such an assertion depends on the fact that assertion carries the conversational implicature that in 'asserting *P* positively you imply, though you don't assert, that you know that *P*' (Moore 1962, p. 277; see also Unger 1975, pp. 255-65). To rephrase Williamson, in performing the speech act of assertion, one represents oneself as having the authority to make that very assertion. If something weaker than knowledge warranted assertion then I could assert '*P* but I don't know that *P*' with impunity.

If the constitutive rule model validates the slogan *only knowledge warrants assertion*, then what does this model have to say about the speech act of supposition? Supposition (or assumption) already has a generally agreed slogan to the effect that *all assumptions are for*

free. Indeed, the validity of the *Rule of Assumptions* in Gentzen style sequent-calculi presupposes the veracity of this slogan. With respect to the Rule of Assumptions, Lemmon (1965, p. 8) has said that

> This rule permits us to introduce at *any* stage of an argument *any* proposition we choose as an assumption of the argument. [...] It may seem dangerously liberal that the rule of assumptions imposes no limit on the kind of assumptions we may make (in particular there is no question of ensuring that assumptions made are true). This is best understood by reminding ourselves that the logician's concern is with the [validity] of the argument rather than the truth or falsity of any assumptions made; hence [this rule] allows us to make any assumptions we please—the job of the logician is to make sure that any conclusion based on them is validly based, *not* to investigate their credentials.

These remarks are instructive. Lemmon is of course right to stress that a speaker can legitimately suppose a sentence which is false, for the logician's business is first and foremost to establish what follows from what. In proving the law of non-contradiction, for example, we must indeed first suppose a sentence which is *necessarily* false. It is a further issue whether in supposing P the logician is free to ignore P's *non*-truth-theoretic credentials. The matter is of course clearer in the case of the *sentential* as opposed to the *propositional* calculus. Propositions just are *bona fide* supposable truth-bearers by default, so goes the thought, while declarative sentences may fail to say that something is the case. Lemmon, I'm sure, would have agreed that the Rule of Assumptions in the sentential calculus would require restriction in order to accommodate those declarative sentences which, for whatever reason, fail to express propositions. In *sentential* logic we thus need to state the Rule of Assumptions as follows: we are permitted to introduce at any stage in a proof any declarative sentence we choose as a premise of the argument only if that sentence is *supposition-apt*. In sequent calculus form this rule is to be given as follows:

Rule of Assumptions. _____ (Provided the sentence Σ is supposition-apt)

$\Sigma \vdash \Sigma$

That is, from the 'null sequent' (or the empty sequent $...\vdash...$) we can infer the sequent that $\Sigma \vdash \Sigma$ only if the credentials of Σ are in order—that, is only if Σ is supposition-apt. Should we wish to work exclusively in a propositional sequent calculus then worries over sentences

156

which fail to express propositions can be put aside. However, the possibility of the propositional version of the liar paradox given above, together with the key thesis of this chapter—that liar sentences/propositions are meaningful but fail to be supposition-apt— means that we need to keep this restriction in place even when stating the rule of assumptions with respect to reasoning with propositions. In order to have a general theory of suppositional inaptitude, however, it is necessary to persevere with the sentential calculus.

The Rule of Assumptions as stated above effectively incorporates into our proof-theory the following constitutive rule governing suppositions:

The S-rule   One must: suppose that *P* only if *P* is supposition-apt.[22]

The question now arises: have we stated this rule strongly enough? In making any speech act one represents oneself to have the authority to do so. In supposing *P*, one represents oneself as *knowing* that *P* is supposition-apt, as knowing that *P* is fit to enable the telos of supposition to be satisfied. But this suggests that the following stronger rule is in fact correct:

The S-rule*   One must: suppose that *P* only if one *knows* that *P* is supposition-apt.[23]

In supposing *P*, a speaker implies, but does not assert, that she knows that *P* is supposition-apt. Arguably, it would be an (albeit artificial) Moorean paradox to suppose the sentence '*P*' while simultaneously cancelling the implicature that one has the authority to make this supposition by asserting (perhaps using the convention of holding up a certain flag) 'I do not know that *P* is supposition-apt'. One does something wrong in simultaneously supposing a sentence but disavowing any knowledge that this sentence is fit to satisfy the telos of supposition. If something less than knowledge that *P* is supposition-apt made supposing *P*

---

[22] As stated, the condition C(*P*) given here is essential but not *unique* to supposition. To ensure that this rule essentially and uniquely governs the act of supposition we also need to build into this rule the proviso that it is not necessary that *P* be true or even warranted, or indeed that *P* be believed by the speaker. Since this proviso plays no part in the discussion which follows I will omit its further mention.

[23] This knowledge requirement might be thought to be too strong. But there is actually no reason to demand that knowledge of supposition-aptness should take the form of a rigorous proof, or even be reflectively accessible to the subject. On the Williamsonian model of assertion, for example, one can satisfy the *A-rule* without knowing that one has done so. The point carries over to issues connected with supposing contingent liar sentences. While the postcard paradox, and analogues, are easily recognisable as such, most contingent liar paradoxes lurk unseen. In general however, even in the absence of any overt evidence of the supposition-aptness of *P*, one suspects that it is not an *easy* possibility that one be mistaken about the supposition-aptness of *P*. Think of the assertion at time t1 of 'what you, the reader, judged to be the case at time t2 is false', where you asserted at time t2 the sentence 'what Patrick judged to be true at t1 is true'. It's not an easy possibility that I should form the

permissible then one could enact the Moorean paradox with impunity, but one cannot. The *S*-rule* thus encodes the injunction: Don't suppose *P* if you don't know that *P* is supposition-apt. In other words, we have to replace the slogan *all assumptions are for free* with the slogan that only knowledge that *P* is supposition-apt permits the supposition that *P*. Thus it would appear that the stronger *S*-rule* is the correct constitutive rule governing supposition. Accordingly, it seems we must modify the Rule of Assumptions to the effect that: we are permitted to introduce at any stage in a proof any declarative sentence we choose as a premise of the argument only if that sentence is *known* to be *supposition-apt*. Thus:

Rule of Assumptions. * _____ (Provided the sentence Σ is *known* to be supposition-apt)

$$\Sigma \vdash \Sigma$$

It now ought to be clear where we go wrong in the paradoxical derivation. One breaks the constitutive *S*-rule* (and indeed the *S*-rule) in supposing liar sentences to be true or in supposing them to be not-true. This is, first and foremost, a mistake at the level of speech acts—a pragmatic mistake. The revised Rule of Assumptions* represents a way of accommodating the possibility of this pragmatic mistake into our proof theory. Once this accommodation is made, then we are in a position to say that in attempting to truth-evaluate the liar sentence we go wrong at the very first step in applying the Rule of Assumptions* to this sentence. Since this rule is indeed a rule of proof, it is the reasoning which is at fault in the liar paradox. The suppositional credentials of liar sentences are such that it is illegitimate to suppose them: the lesson of the liar is that not all assumptions are for free.[24] On that basis, we have found a principled reason to refuse to suppose that L is true and a principled reason to refuse to suppose that L is not true. But are matters really so straightforward?

---

false belief that my assertion is supposition-apt. Where I lack evidence of the supposition-aptness of *P*, the world remains, in general, supposition-friendly.

[24] Of course not all deductive systems contain the Rule of Assumptions—axiomatic presentations being the most notable example. Since for every proper axiomatic presentation of the sentential calculus there is a corresponding natural deduction presentation then this ought to be no obstacle to the full generality of this proposal. In this respect, consider the simplest template of the liar paradox whereby a substitution of Tarski's T-schema is 'L is not true' is true iff L is not true, and where by substitution this yields the paradoxical biconditional L is true iff and L is not true. Since Tarski's T-schema is a derived principle requiring the rules of truth-introduction and truth-elimination, together with conditional proof, and the rule of assumptions, then this simple template presents no special problem. To restrict the rule of assumptions to non-liar sentences is just to

The Standard Solution has, throughout its various guises, been a conspicuous failure owing to the problem of the Strengthened liar paradox. Is there a form of strengthened liar sentence which might regenerate the paradox for this version of the Standard Solution? Any sentence which says of itself that it is not supposition-apt seems to be a good candidate for a strengthened liar sentence for this proposal. Let the logical form of this candidate strengthened liar sentence be represented by the equality SL = 'SL is not supposition-apt' and suppose for the sake of argument that

(1)   SL is not supposition-apt

by the rule of truth-introduction we infer

(2)   'SL is not supposition-apt' is true

and by substitution we derive

(3)   SL is true.

If we allow that it is sufficient for SL to be supposition-apt that it be true then we can further infer

(4)   SL is supposition-apt

which contradicts (1) and so by negation-introduction we infer

(5)   ~(SL is not supposition-apt)

which is just demonstrate that the sentence which says of itself that it is not supposition-apt is false—this sentence *is* in fact supposition-apt (since it is false) contrary to what it says of itself. (Cf. The sentence which says of itself that it contains ten words.) Unlike other forms of

---

restrict Tarski's T-schema likewise, providing just the sort of principled constructive restriction of the T-schema

the Standard Solution the proposal in hand does not appear to regenerate the paradox in some refined form. This provides a strong *prima facie* reason for thinking that a proposal of this sort is along the right lines. Is there a 'revenge problem' for this proposal?

In the literature on the liar, the revenge problem (for whatever proposal in hand) has in general been confused with the problem of the strengthened liar paradox. Roughly, a solution suffers from the revenge problem when it has pathological but not necessarily inconsistent consequences (see fn. 10). It would appear that there is a form of revenge problem for the solution proposed here, and it can be framed as follows: in order to test whether a sentence passes the supposition test one must first suppose this sentence in order to demonstrate what its (putative) logical consequences might be. If a sentence fails to pass the test then it is illegitimate to suppose it—but that was just what we needed to do in order to test its credentials via the supposition test. Briefly put, we seem to be supposing the liar sentence in order to show that it is not legitimately supposable, but if it is not legitimately supposable then we are not entitled to suppose it *tout court*. Even though this revenge problem would appear to be a pragmatic rather than a logical paradox, it nonetheless demands a response.

One thought might be that the *S*-rule* (and indeed the *S*-rule) are stated too strongly. Independently of any worries concerning the liar paradox, there are grounds to think that both these rules are too prohibitive in any case. Consider the possibility that a certain very complex (but grammatical) sentence Ø encodes a category mistake. Such a sentence fails to say that something is the case. Suppose that a speaker nonetheless believes Ø to be supposition-apt and performs the speech act of supposing Ø in order to see what logically follows. Given the *S*-rule and the *S*-rule*, this speaker has done something wrong in supposing Ø to be true. Yet there may be no other way of revealing that Ø encodes a category mistake other than by supposing it to be true/not-true and applying rules of inference to its sentential and sub-sentential structure. In response to this possibility, one might try to considerably weaken the *S*-rule as follows:

The S-rule**   One must: suppose that *P* only if one does not know that *P* fails to be supposition-apt.

This weaker rule encodes the injunction: don't suppose *P* if you know *P* fails to be supposition-apt. Thus, for instance, in the absence of any warrant for believing *P* to be

---

(and the equivalence schema) that Horwich (1998a, p. 42) hopes for (see fn. 6 above).

essentially unfit for suppositional reasoning we are free to suppose P with impunity. But this weaker rule just seems too liberal—surely there ought to be something impermissible about supposing meaningless sentences to be true! A better response to this problem is to distinguish between the primary and secondary goals of supposition.

The primary teleological norm governing suppositions is the norm distinguished hitherto: in supposing P, in order to see if some sentence Q logically follows, aim to have a warrant either to deny or accept that $\vdash P \rightarrow Q$. In supposing the sentence $\emptyset$ to be true one will indeed essentially fail to satisfy this primary goal of supposition. Nonetheless, in such cases a secondary norm may legitimately come into force, namely: in supposing P, aim to have a warrant to either accept or deny the thesis that P is supposition-apt. In such cases, the rules of the game of supposition have changed and a speaker is accordingly permitted to suppose P at least insofar as they are now aiming to satisfy a different goal—the goal of testing the suppositional credentials of P. It seems we can give a constitutive rule to accommodate this secondary teleological norm as follows:

The P-rule.   One must: suppose that P only if one has presupposed that P is
            supposition-apt.

This rule looks rather cumbersome, but its effect is to permit a speaker to suppose sentences which are supposition-inapt, and indeed which might be *known* to be supposition-inapt. Consider again the complex category mistake $\emptyset$. Suppose I know that $\emptyset$ is a category mistake but I want to communicate this fact to others by supposing it to be true and then subjecting it to certain rule of inference in order to reveal its pathological nature. In this case, not only does one break both the *S-rule* and *S-rule** in supposing $\emptyset$, but also the weaker *S-rule***. The *P-rule*, on the other hand, is not broken. This rule encodes the weak injunction: do not suppose P if you have not firstly presupposed that P is supposition-apt. One represents oneself to have the authority to suppose sentences which are potentially, actually, or actually known to be supposition-inapt simply on the basis that one undertakes a commitment to discharge the presupposition that these sentences are supposition-apt if these sentences reveal (or re-reveal) themselves to be essentially unfit for suppositional reasoning.

On the plausible assumption that a speech act can be identified by the teleological and constitutive rules which uniquely and essentially govern that act, then we can thus distinguish two species of the speech act of supposition: supposing P in order to see what the logical

consequences of this sentence are, and supposing P in order to establish whether P is supposition-apt. The thought now goes that just as we can legitimately suppose the sentence Ø, at least insofar as we are bound by a teleological norm to uncover the suppositional credentials of Ø, we can likewise legitimately suppose that L is true in order to test the suppositional credentials of L. If this is right, then there is no genuine revenge problem for a view of this sort.

5.9 Semantic closure

Lastly, we may ask if the proposed solution is able to meet the requirements of semantic closure? The aim of the solution developed here is to indeed show that a semantically closed language can after all be consistent once we appropriately restrict the Rule of Assumptions. But we must take care what is meant by semantic closure. For Tarski, a language is semantically closed when

> the language in which the antinomy is constructed contains, in addition to its
> expressions, also the names of these expressions, as well as semantic terms such
> as the term '*true*' referring to sentences of this language; we have also assumed
> that all sentences which determine the adequate usage of this term can be asserted
> in the language.[25]

Others, such as Herzberger (1970, p. 26), go much further, and argue that a semantically closed language should at the very least contain 'the *means* for recording the truth-value of each of its own sentences' (my italics). The theory just offered is semantically closed in Tarski's sense, but not strictly in Herzberger's. Merely for the language to contain its own truth-predicate does not in itself entail that the language contains the expressive means for assigning truth-values to its sentences. On the proposed theory, a speaker is essentially prevented from assigning truth-values to liar sentences for it is improper to suppose that such sentences have truth-values—a claim which falls short of asserting that they lack truth-values. This might seem an unwelcome consequence. As Parsons (1984, pp. 148-9) notes, it is 'commonly said, regarding the paradoxes, that one must "buy consistency at the price of expressive completeness"'. But this is a trade-off the deflationist is perfectly willing to pay. Given that on a deflationary view, the semantics of the language is not given by a truth-

---

[25] Tarski (1944) in Linsky (1952, p. 20).

conditional theory of content but by a theory in which truth plays no explanatory role, then it is grist for the deflationist's mill that the theory of truth is expressively incomplete (see fn. 12). The real expressive adequacy of a semantics for the language must reside at the level of the sentences which, to paraphrase Tarski, determine the adequate usage of the terms in the language. The sentence 'L is not supposition-apt' is a sentence of just this sort—and we have seen that the language can contain such sentences without fear of generating further paradox. If this is correct, then truth does not and should not play any explanatory role in the dissolution of the liar paradox.

BIBLIOGRAPHY

Akiba, K. (1999): 'On super- and subvaluationism: a classicist's reply to Hyde', *Mind*, 108, pp. 727–32.

Alston, W. P. (1993): 'Epistemic desiderata', *Philosophical and Phenomenological Research*, 53, pp. 527–551.

Armour-Garb, B. (forthcoming): 'Deflationism and Dialetheism'.

Armstrong, D. M. (1973): *Belief, Truth, and Knowledge*, Cambridge: Cambridge University Press.

Austin, J. L. (1946): 'Other Minds', *Proceedings of the Aristotelian Society*, Suppl. Vol. 20, pp. 148–87, reprinted in S. Bernecker and F. Dretske (eds), *Knowledge: Readings in Contemporary Epistemology*, Oxford: Oxford University Press, 2000, pp. 339–346.

Ayer, A. J. (1935): 'The criterion of truth', *Analysis*, 3.

Bach, K. (1985): 'A rationale for reliabilism', *The Monist*, 68, pp. 246–63.

Bar-Hillel, Y. (1957): 'New light on the liar', *Analysis*, 18, pp. 1–6.

Black, M. (1939): 'Vagueness: an exercise in logical analysis', *Philosophy of Science*, 4, pp. 427–55.

Bochvar, D. A. (1939): 'On a three-valued logical calculus and its application to the analysis of contradictories', *Metematiceskij sbornik*, 4.

Bonjour, L. (1985): *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

Bonjour, L. (1992): 'Externalism/internalism', in J. Dancy and E. Sosa (eds), *A Companion to Epistemology*, Oxford: Blackwells, 1992, pp. 132–36.

Brandom, R. (1998): 'Insights and blindspots of reliabilism', *Monist*, 81, pp. 371–92.

Burgess, J. (1990): 'The sorites paradox and higher-order vagueness', *Synthese* 85, pp. 417–74.

Burgess, J. (1998): 'In defence of an indeterminist theory of vagueness', *Monist*, 81, pp. 233–52.

Cargile, J. (typescript): 'Supposing for the sake of argument', July 1999.

Channell, J. (1994): *Vague Language*, Oxford: Oxford University Press.

Chellas, B. F. (1980): *Modal Logic: An Introduction*, Cambridge: Cambridge University Press.

Chisholm, R. (1966): *Theory of Knowledge*, Englewood Cliffs, NJ: Prentice-Hall.

Cohen, S. (1988): 'How to be a fallibilist', *Philosophical Perspectives*, 2, pp. 91–123.

Conee, E. and Feldman, R. (1998): 'The generality problem for reliabilism', *Philosophical Sudies*, 89, pp. 1–29.

DeRose, K, (1995): 'Solving the sceptical problem', *Philosophical Review*, 104, pp. 1–52.

Dretske: F. (1970): 'Epistemic operators', *Journal of Philosophy*, 67, pp. 1007–23.

Dretske: F. (1971): 'Conclusive reasons', *Australasian Journal of Philosophy*, 49, pp. 1-22.

Dummett, M. A. E. (1973): *Frege: Philosophy of Language*, London: Duckworth.

Dummett, M. A. E. (1975): 'Wang's paradox', *Synthese*, 30, pp. 301–24.

Dummett, M. A. E. (1976): 'What is a theory of meaning (II)' in G. Evans and J. McDowell (eds), *Truth and Meaning*, Oxford: Oxford University Press.

Dummett. M. A. E. (1978): *Truth and Other Enigmas*, London: Duckworth.

Dummett, M. A. E. (1991): *The Logical Basis of Metaphysics*, London: Duckworth.

Dummett, M. A. E. (1993): *The Seas of Language*, Oxford: Oxford University Press.

Edgington, D. (1993): 'Wright and Sainsbury on higher-order vagueness', *Analysis*, 53, pp. 193–200.

Edgington, D. (1996): 'Vagueness by degrees', in R. Keefe and P. Smith (eds), *Vagueness: A Reader*, pp. 294–316.

Feldman (1974): 'An alleged defect in Gettier counter-examples', *Australasian Journal of Philosophy*, 52, pp. 68–9.

Feldman, R. (1981): 'Fallibilism and knowing that one knows', *Philosophical Review*, 110, pp. 266–82.

Field, H. (1992): Critical Notice of *Truth* by P. Horwich, *Philosophy of Science*, 59, pp. 321–330.

Field, H. (1994a): 'Deflationist views of meaning and content', *Mind*, 103, pp. 247–285.

Field, H. (1994b): 'Disquotational truth and factually defective discourse', *Philosophical Review*, 103, pp. 405–452.

Field, H. (2001): *Truth and the Absence of Fact*, New York: Oxford University Press.

Fine, K. (1975): 'Vagueness, truth, and logic', *Synthese*, 30, pp. 265–300.

Fraassen, B. C. van (1966): 'Singular terms, truth-value gaps, and free logic', *Journal of Philosophical Logic*, 63, pp. 481–495.

Fraassen, B. C. van (1968): 'Presupposition, implication, and self-reference', *Journal of Philosophy*, 65, pp. 136-52.

Frege, G. (1903): *Grundgesetze der Arithmetik, Begriffsschriftlich Abgeleitet, Volume II*, Jena: Hermann Pohle.

Frege, G. (1918): 'The thought: a logical inquiry', *Mind*, 65, pp. 289–311.

Frege, G. (1979): *Posthumous Writings*, Oxford: Blackwells.

Gettier, E. (1963): 'Is justified true belief knowledge?', *Analysis*, 23, pp. 121–3.

Ginet, C. (1970): 'What must be added to knowing to obtain knowing that one knows', *Synthese*, 21, pp. 163–86.

Goldman, A. (1967): 'A causal theory of knowing', *Journal of Philosophy* 64, pp. 357–72.

Goldman, A. (1976): 'Discrimination and perceptual knowledge', *Journal of Philosophy*, 73, pp. 771–91.

Goldman, A. (1986): *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.

Green, M. S. (2000): 'The Status of Supposition', *Noûs*, 34, pp. 376–99.

Guttenplan, S. (1994): 'First-person authority', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwells, 1994, p. 291.

Haack, S. (1978): *Philosophy of Logics*, Cambridge: Cambridge University Press.

Hardin, C. (1990): 'Colour and illusion', in W. G. Lycan (ed.), *Mind and Cognition: A Reader*, Oxford: Blackwells, pp. 555–567.

Harman, G. (1973): *Thought*, Princeton: Princeton University Press.

Heck, R. (1993): 'A note on the logic of (higher-order) vagueness', *Analysis*, 53, pp. 201–8.

Herzberger, H. (1970): 'Truth and modality in semantically closed languages', in R. L. Martin (ed.), *The Paradox of the Liar*, New Haven: Yale University Press.

Hetherington, S. (1998): 'Actually knowing', *Philosophical Quarterly*, 48, pp. 453–469.

Hetherington, S. (1999): 'Knowing failably', *Journal of Philosophy*, Vol. XCVI, pp. 565–587.

Hintikka, J. (1962): *Knowledge and Belief*, Ithaca, NY: Cornell University Press.

Holton, R. (2000): 'Minimalism and truth-value gaps', *Philosophical Studies*, 97, pp. 135–65.

Horwich, P. (1990): *Truth*, 1st edition, Oxford: Blackwells.

Horwich, P. (1998a): *Truth*, 2nd edition, Oxford: Oxford University Press.

Horwich, P. (1998b): *Meaning*, Oxford: Oxford University Press.

Horwich, P. (1997): 'The nature of vagueness', *Philosophical and Phenomenological Research*, 57, pp. 929–35.

Hugly, P and Sayward, C. (1992): 'Redundant truth', *Ratio* (new series) V, pp. 24–37.

Hyde, D. (1994): 'Why higher-order vagueness is a pseudo-problem', *Mind*, 103, pp. 35–41.

Hyde, D. (1997): 'From heaps of gaps to heaps of gluts', *Mind*, 106, pp. 641–60.

Kalderon, M. (1997): 'The transparency property of truth', *Mind*, 106, pp. 475–97.

Kamp, H. (1981): 'The paradox of the heap', in U. Monnich (ed.), *Aspects of Philosophical Logic*, Dordrecht: Reidel.

Kearns, J. T. (1997): 'Propositional logic of supposition and assertion', *Notre Dame Journal of Formal Logic*, 38, pp. 325–349.

Kearns, J. T. (forthcoming): 'A full system of illocutionary propositional logic'.

Keefe, R. and Smith, P. (eds) (1996): *Vagueness: A reader*, Cambridge, MA: MIT press.

Keefe, R. (2000): *Theories of Vagueness*, Cambridge: Cambridge University Press.

Kirkham, R. (1992): *Theories of Truth*, Cambridge, MA: MIT Press.

Kitcher, P. (2000): '*A priori* knowledge revisited', in P. Boghossian and C. Peacocke (eds), *New Essays on the A Priori*, Oxford: Oxford University Press, pp. 65–91.

Kleene, S. C. (1952): *Introduction to Metamathematics*, North Holland.

Koons, R. (1994): 'A new solution to the sorites problem', *Mind* 103, pp. 439–449.

Lehrer, K. (2000): *Theory of Knowledge*, 2nd edition, Boulder, CO: Westview Press.

Lemmon, E. J. (1965): *Beginning Logic*, London: Thomas Nelson and Sons.

Lewis, D. (1973): 'Causation', *Journal of Philosophy*, 70, pp. 556–67.

Lewis, D. (1996): 'Elusive knowledge', *Australasian Journal of Philosophy*, 74, pp. 549–67.

Linsky, L. (ed.) (1952): *Semantics and the Philosophy of Language*, Urbana: University of Illinois Press.

Lukasiewicz, J. (1930): 'Many-valued systems of propositional logic', in S. McCall (ed.), *Polish Logic*, Oxford: Oxford University Press, 1967.

Luper-Foy, S (1987a): 'The possibility of scepticism', in S. Luper-Foy (1987b), pp. 219–41.

Luper-Foy, S. (ed.) (1987b): *The Possibility Of Knowledge: Nozick and his Critics*, Totowa, NJ: Rowman and Littlefield.

Machina, K. (1976): 'Truth, belief, and vagueness', *Journal of Philosophical Logic*, 5, pp. 47–78.

Mackie, J. L. (1973): *Truth, Probability and Paradox*, Oxford: Clarendon Press.

Mackie, J. L. (1977): *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.

McCall, S. (1970): 'A non-classical theory of truth with an application to intuitionism', *American Philosophical Quarterly*, 7, pp. 83–88.

McDowell, J. (1976): 'Truth-conditions, bivalence, and verificationism', in G. Evans and J. McDowell (eds), *Truth and Meaning*, Oxford: Oxford University Press, pp. 42-66.

McGee, V. (1989): 'Applying Kripke's theory of truth', *Journal of Philosophy*, 86, pp. 530–539.

McGee, V. (1991): *Truth, Vagueness, and Paradox*, Indianapolis: Hackett.

McGee, V. and McLaughlin, B. (1995): 'Distinctions without a difference', *Southern Journal of Philosophy* 33, (supplement), pp. 203–51.

Merricks, T. (1995): 'Warrant entails truth', *Philosophical and Phenomenological Research*, 50, pp. 841–855.

Mills, A. (1995): 'Unsettled problems with vague truth', *Canadian Journal of Philosophy*, 25, pp. 103–117.

Moore, G. E. (1962): *Commonplace Book: 1919–1953*, London: Allen and Unwin.

Morton, A. (2000): 'Saving epistemology from the epistemologists: recent work in the theory of knowledge', *British Journal for the Philosophy of Science*, 51, special supplement, pp. 685–704.

Nozick, R. (1981): *Philosophical Explanations*, Cambridge, MA: Harvard University Press, and reprinted in S. Luper-Foy (1987b), as 'Knowledge and scepticism', pp. 19-115.

O'Leary-Hawthorne, J. and Oppy, G. (1997): 'Minimalism and truth', *Noûs*, 31, pp.170–96.

Parsons, T. (1984): 'Assertion, denial, and the liar paradox', *Journal of Philosophical Logic*, 13, pp. 137–52.

Peirce, C. S. (1902): 'Vague', in J. M. Baldwin (ed.), *Dictionary of Philosophy and*

Psychology, New York: Macmillan, p. 748.

Plantinga, A. (1993): *Warrant and Proper Function*, Oxford: Oxford University Press.

Plantinga, A. (1995): 'Précis of *Warrant: The Current Debate* and *Warrant and Proper Function*', *Philosophical and Phenomenological Research*, 55, pp. 393–96.

Pollock, J. L. (1986): *Contemporary Theories of knowledge*, Lanham, MD: Rowman and Littlefield.

Prawitz, D. (1965): *Natural Deduction: A Proof-Theoretical Study*, Uppsala: Almqvist and Wiksells.

Prichard, H. A. (1950): *Knowledge and Perception*, Oxford: Clarendon Press.

Priest, G. (1987): *In Contradiction: A Study of the Transconsistent*, The Hague: Nijhoff.

Priest, G. (1994): Review of *Truth, Vagueness, and Paradox* by Vann McGee, *Mind*, 103, pp. 387–391.

Priest, G. (2000): 'Truth and contradiction', *Philosophical Quarterly*, 50, pp. 305–19.

Putnam, H. (1983): 'Vagueness and alternative logic', *Erkenntnis*, 19, pp. 297-314.

Quine, W. V. O. (1970): *Philosophy of Logic*, Englewood Cliffs: Prentice-Hall.

Ramsey, F. (1927): 'Facts and propositions', *Proceedings of the Aristotelian Society*, Vol. 7., pp. 153–70.

Reed, B. (2000): 'Accidental truth and accidental justification', *Philosophical Quarterly*, 50, pp. 57–67.

Rescher, N. (1969): *Many-Valued Logic*, New York: McGraw-Hill.

Russell, B. (1923): 'Vagueness', *Australasian Journal of Philosophy and Psychology*, 1, pp. 84–92.

Sainsbury, M. (1989a): 'Tolerating vagueness', *Proceedings of the Aristotelian Society*, 89, pp. 33–48.

Sainsbury, M. (1989b): 'What is a vague object?', *Analysis*, 49, pp. 99–103.

Sainsbury, M. (1990): 'Concepts without boundaries', inaugural lecture, published by King's College London, Dept. of Philosophy.

Sainsbury, M. (1991): 'Is there higher-order vagueness?', *Philosophical Quarterly*, 41, pp. 167–82.

Sainsbury, M. (1995a): 'Vagueness, ignorance, and margin for error', *British Journal for the Philosophy of Science*, 46, pp. 589–601.

Sainsbury, M. (1995b): 'Why the world cannot be vague', *Southern Journal of Philosophy*, 33, (Supplement), pp. 63–81.

Sainsbury, M. (1997): 'Easy possibilities', *Philosophical and Phenomenological Research*, 57, pp. 907–19.

Sainsbury, M and Williamson, T. (): 'Sorites', in B. Hale and C. Wright, *A Companion to the Philosophy of Language*, Oxford: Blackwells, pp. 458–84.

Sanford, D. (1976): 'Competing semantics of vagueness: many values versus super-truth', *Synthese* 33, pp. 195–210.

Shapiro, S. (1997): *The Philosophy of Mathematics: Structure and Ontology*, New York: Oxford University Press.

Shoemaker, S. (1988): 'On knowing one's own mind, *Philosophical Perspectives*, 2, pp. 183–209.

Shoemaker, S. (1994): 'Introspection', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Oxford: Blackwell Publishers, 1994, pp. 395–400.

Shope, R. K. (1983): *The Analysis of Knowing: A Decade Research*, Princeton: Princeton University Press.

Simmons, K. (1999): 'Deflationary truth and the liar', *Journal of Philosophical Logic*, 28, pp. 455–488.

Simons, P. (1992): 'Vagueness and ignorance', *Aristotelian Society*, suppl. 66, pp. 163–77.

Skyrms, B. (1970): 'Return of the liar: three-valued logic and the concept of truth', *American Philosophical Quarterly*, 7, pp. 153–161.

Smiley, T. (1960): 'Sense without denotation', *Analysis*, 20, pp. 125–35.

Smith, J. W. (1984): 'The surprise examination on the paradox of the heap', *Philosophical Papers*, 13, pp. 43–56.

Soames, S. (1999): *Understanding Truth*, Oxford/New York: Oxford University Press.

Sorenson, R. A. (1985): 'An argument for the vagueness of "vague"', *Analysis*, 45, pp. 134–7.

Sorenson, R. A. (1988): *Blindspots*, Oxford: Oxford University Press.

Sorenson, R. A. (1989): 'The ambiguity of vagueness and precision', *Pacific Philosophical Quarterly* 70, pp. 174–83.

Sosa, E. (1996): 'Postscript to "Proper functioning and virtue epistemology"', in J. L. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honour of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman and Littlefield.

Sosa, E. (1999): 'How to defeat opposition to Moore', in J. Tomberlin (ed.), *Philosophical Perspectives 13: Epistemology*, Oxford: Blackwell Publishers, pp. 141–53.

Sosa, E. (2000): 'Scepticism and contextualism', in J. Tomberlin (ed.), *Philosophical Issues 10, Noûs* suppl. issue 34, pp. 1–42.

Stirton, W. R. (1997): 'Anti-realism, truth-conditions, and verificationism', *Mind*, 106, pp. 697–716.

Stroud, B. (1994): 'Scepticism, "externalism", and the goal of epistemology', *Proceeding of the Aristotelian Society, Supplementary Volume*, reprinted in his *Understanding Human Knowledge*, New York: Oxford University Press, 2000.

Tarski, A. (1944): 'The semantic conception of truth', *Philosophical and Phenomenological Research*, 4, pp. 341–76, and in Linsky (1952).

Tennant, N. (1997): *The Taming of the True*, Oxford: Oxford University Press.

Travis, C. (1999): 'Sublunary intuitionism', in Brandl and Sullivan (eds), *New Essays on the Philosophy of Michael Dummett*, Rodopi: BV Editions.

Tye, M. (1990): 'Vague objects', *Mind*, 99, pp. 535–57.

Tye, M. (1994): 'Sorites paradoxes and the semantics of vagueness', *Philosophical Perspectives, 8: Logic and Language*, in J. E. Tomberlin (ed.), Atascadero, CA: Ridgeview, pp. 281–93.

Tye, M. (1995): 'Vagueness: welcome to the quicksand', *Southern Journal of Philosophy*, 33, (supplement), pp. 1–22.

Unger, P. (1968): 'An analysis of factual knowledge', *Journal of Philosophy* 65, pp. 157–70.

Unger, P. (1975): *Ignorance: a Case for Scepticism*, Oxford: Oxford University Press.

Unger, P. (1979): 'There are no ordinary things', *Synthese*, 41, pp. 117–54.

Weiss, S. (1976): 'The sorites fallacy: what difference does a peanut make?', *Synthese*, 33, pp. 253–72.

Wheeler, S. C. (1979): 'On that which is not', *Synthese*, 41, pp. 155–94.

Williams, M. (1996): *Unnatural Doubts: Epistemological Realism and the Basis for Scepticism*, Princeton: Princeton University Press.

Williamson, T. (1992a): 'Inexact knowledge', *Mind*, 101, pp. 217–42.

Williamson, T. (1992b): 'Vagueness and ignorance', *Proceedings of the Aristotelian Society, Supplementary volume*, 66, pp. 145–62.

Williamson, T. (1994): *Vagueness*, London/New York: Routledge.

Williamson, T. (1995): 'Definiteness and knowability', *Southern Journal of Philosophy*, 33, (supplement), pp. 171–91.

Williamson, T. (1996a): 'Cognitive homelessness', *Journal of Philosophy*, 93, pp. 554–73.

Williamson (1996b): 'Wright on the epistemic conception of vagueness', *Analysis*, pp. 39–45.

Williamson, T. (1996c): 'Knowing and asserting', *Philosophical Review*, 105, pp. 489–523.

Williamson, T. (1996d): 'What makes it a heap?', *Erkenntnis*, 44, pp. 327–339.

Williamson, T. (1997a): 'Imagination, stipulation, and vagueness', in E. Villanueva (ed.) *Truth: Philosophical Issues*, vol. 8, Atascadero, CA: Ridgeview.

Williamson, T. (1997b): 'Unreflective realism', *Philosophical and Phenomenological Research*, 56, pp. 905–09.

Williamson, T. (1999): 'On the structure of higher-order vagueness', *Mind*, 108, pp. 127–43.

Williamson, T. (2000): *Knowledge and its Limits*, Oxford: Oxford University Press.

Wright, C : (1975): 'On the coherence of vague predicates', *Synthese*, 30, pp. 325–65.

Wright, C: (1976): 'Language mastery and the sorites paradox', in G. Evans and J. McDowell (eds), *Truth and Meaning: Essays in Semantics*, Oxford: Clarendon Press, pp. 223–47.

Wright, C. (1987): 'Further reflections on the sorites paradox', *Philosophical Topics*, 15,

pp. 227–90.

Wright, C. (1992a): *Truth and Objectivity*, Cambridge, MA: Harvard University Press.

Wright, C. (1992b): 'Is higher-order vagueness coherent?', *Analysis*, 52, pp. 129–39

Wright, C. (1995): 'The epistemic conception of vagueness', *Southern Journal of Philosophy*, 33 (supplement), pp. 133–59.

Wright, C. (1999): 'Truth: a traditional debate reviewed', *Canadian Journal of Philosophy*, suppl. vol. 24, and reprinted in S. Blackburn and K. Simmons (eds), *Truth*, Oxford: Oxford University Press, 1999, pp. 203–38.

Wright, C. (2001): 'On being in a quandary', *Mind*, 110, pp. 45–97.

Wright, G. H. von (1986): 'Truth, negation, and contradiction', *Synthese*, 66, pp. 3–14.

Wright, G. H. von (1987): 'Truth-Logics', *Logique et Analyse*, NS 30, pp. 311–34.

Zagzebski, L. (1994): 'The inescapability of Gettier problems', *Philosophical Quarterly*, 44, pp. 65–73.

Zagzebski, L. (1999): 'What is knowledge?', in J. Greco and E. Sosa (eds), *The Blackwells Guide to Epistemology*, Oxford: Blackwells, 1999, pp. 92–116.

Ziff, P. (1984): *Epistemic Analysis*, Dordrecht: Reidel.